

Better to be talked about...

The prominence of embryonic stem cells as a key issue in the US presidential election campaign is, at best, a mixed blessing for science.

The two main candidates for the world's most powerful elective office are poles apart on several science-related issues. Their answers to a series of questions on these topics reveal a rich tapestry of disagreement (see News Feature, page 238).

George W. Bush advocates the modernizing of America's nuclear-weapons stockpile, whereas John Kerry would halt such work. Kerry states emphatically that climate change requires "concrete steps" to limit emissions of greenhouse gases in the United States, and to restart international negotiations on emissions limits; Bush acknowledges that climate change is a "serious long-term issue" but says that "considerable uncertainty" surrounds its impact. Kerry promises to clean up the process from which the government gets scientific advice, insinuating that Bush has sullied it. And so on.

But for the casual observer (and anyone who hasn't made up their mind yet is probably observing pretty casually), there is only one scientific issue in the election: human embryonic stem-cell research.

Bush wanted to defuse the issue when he announced guidelines on 9 August 2001 whereby researchers could obtain federal funding to work with embryonic stem-cell lines derived before then. Kerry recognized the unpopularity of Bush's compromise and made much of it at this year's Democratic convention in Boston, where, one of his acolytes bizarrely boasts, he mentioned embryonic stem cells more often than unemployment. His strategists want to use embryonic stem-cell research as a "wedge issue" in the campaign. In the parlance of the political consultant, a wedge is a block of very hard material of triangular cross-section that will be hammered repeatedly into something softer — one's opponent's support base — in order to split it up.

Some science lobbyists argue (without much conviction) that the debate is a healthy one, exposing the public to some of the intricacies of cell biology. But the emergence of the stem-cell issue as part of this bitterly divisive campaign is problematic for developmental biology, and for science as a whole.

For a start, the argument threatens to subsume all other aspects of biology in the public mind. When one staff member at a scientific society took a vacation last month to Alaska, he told a fellow vacationer that he worked "in biomedical research". "Oh," said the man, a resident of Cleveland, in the great swing state of Ohio. "So you work on stem cells?"

More seriously, the debate lumps the scientific community on one side of what is really an ethical and moral, rather than a scientific, question. There are many scientific questions that could be asked in the campaign to which the scientific method can perhaps provide an answer, such as: "Is global warming being caused primarily by greenhouse-gas emissions?"

The stem-cell question falls into a different category. The question: "In what circumstances is it appropriate to use human embryos for fundamental scientific research?" is essentially a moral and ethical one. Scientists can inform the public on aspects of the question, but they cannot answer it.

Many scientists see Bush's guidelines as an unnecessary barrier to important research, and rightly so. It is entirely appropriate that they should seek to engage the public on the question. But the rallying of various prominent supporters, and of disease-specific lobby groups, to the cause has resulted in some over-the-top claims being made about the likely impact of stem-cell research on treatments for diseases such as Alzheimer's.

As Oscar Wilde might have said, it is better for science to be talked about in the campaign than for it not to be talked about. But the nature of the current conversation carries risks. These can be minimized if scientists refrain from making extravagant public claims on behalf of embryonic stem-cell research, and endeavour, against all the odds, to maintain a genuine dialogue with people of good faith who are worried about the moral and ethical limits of developmental biology. ■

How to interfere with RNA

Using RNA to manipulate gene expression is a powerful experimental tool, but can lead researchers astray.

There are many reasons to celebrate the recent discovery of RNA interference (RNAi). The finding that small double-stranded RNA molecules can silence gene expression has had a huge impact on both biological research and technological applications.

For example, it adds a further layer of complexity to our understanding of gene regulation; we have been alerted to the hundreds of small RNAs with potential regulatory roles in development and physiology; and new pathways and enzymes involved in RNA processing are rapidly emerging. RNAi has been exploited as an experimental tool to inactivate gene function, and has enabled large-scale genetic experiments to be performed more effectively in mammalian cells, when they had hitherto been feasible only in lower organisms. There is therapeutic potential for RNAi too, with the first clinical trials currently being put forward for regulatory approval. The *Nature* Insight in this issue reviews this exciting field.

But it also carries an important note of caution about RNAi as a technology, which is discussed in a review by Gregory Hannon and John Rossi (see page 371). RNAi has many advantages over other methods of gene inactivation in its apparent ease and specificity. But just as with these other methods, the specificity of silencing is not absolute and there is a danger of 'off-target' effects. Hannon and Rossi explain the potential pitfalls and offer some "rules of the road" for effective RNAi experiments.

We fully endorse their recommendations. Good experimental design and controls can alleviate many of the specificity problems, and we will work with peer reviewers to keep abreast of such requirements in this fast-moving field (see also *Nature Cell Biol.* **5**, 489–490; 2003). In the future, a deeper understanding of the mechanisms of RNAi should allow better design of RNAi triggers and enable non-specific effects to at least be predicted, if not eliminated. ■





Deep trouble
Aquarium is all at sea over how to get rid of its tuna
p233



Caught out
Italian minister's support for science backfires
p234



Cold comfort
Beleaguered ski resorts seek lift for business on glaciers
p235



Up, up and...up
Price tag for US Moon mission goes into orbit
p237

Security restrictions lead foreign students to snub US universities

Geoff Brumfiel, Washington

The number of foreign graduate students admitted to top US research institutions has declined precipitously this year, according to data released on 7 September.

Admissions from China, by far the largest source of foreign scientists in the United States, fell by a third this year, compared with 2003, according to the Council of Graduate Schools. The council surveyed 126 institutions, including leading private and public research universities in the United States.

The numbers appear at a time when government officials are pledging to reform a visa system that, most observers agree, has created mounting difficulties for incoming students and researchers since the terrorist attacks of 11 September 2001. And both President George Bush and Democrat candidate John Kerry have told *Nature* that if they win November's election they will take steps to ensure that foreign scientists are made welcome in the United States (see page 238).

The survey showed a 19% decrease in the number of foreign students admitted to graduate programmes in the life sciences and a 17% drop in admissions in the physical and Earth sciences. Admissions from China, India and South Korea, which between them provide the lion's share of foreign graduate students in the United States, were all down sharply (see graph, right).

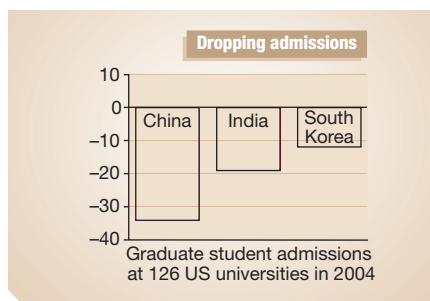
The report seems to back up years of complaints from academics and administrators that US security procedures are hurting the country's ability to attract graduate students from around the world (see *Nature* 427, 190–195; 2004).

The survey, which measures the number of students actually admitted to US schools — as opposed to the number of applications received — is the first concrete evidence that fewer foreign graduate students are coming this year, says Heath Brown, director of research and policy analysis at the Washington-based council.

"I think that it points to a very serious problem," says Peggy Blumenthal, vice-president for educational services at the Institute of International Education in New York. Blumenthal says that new immigration rules



Clamp-down: mounting security controls at US entry points may be deterring foreign graduates.



introduced after the 11 September attacks are partly to blame for the decline. These require foreign scientists to undergo security checks with more than one US agency. The checks, along with required interviews and a new Department of Homeland Security student-tracking system, have caused lengthy delays for some visitors.

Just two days after the study was released, government officials told Congress that they were trying to reduce waiting times. "We have made great progress in improving the inter-agency security-clearance process in recent months by moving from a paper-based system to electronic transmission," said Janice Jacobs, head of visa services at the Department of State.

Stewart Verdery, assistant secretary for border security at the Department of Homeland Security, added that his agency was working with the state department to extend security clearances granted to foreign students and scientists to more than a year. "We are looking at ways to make the visa application process smoother," he said.

Wendy White, head of the international office at the National Academy of Sciences, calls the proposed changes "terrific" but says that she would like to see more details about how and when they will be implemented. She adds that her office is still averaging more than 100 complaints every month from scientists who are experiencing delays.

Tough economic conditions in some states, including California, have led to overall reductions in the number of graduate students admitted at public universities, says Brown, and domestic graduate student enrolment is also down, by 5%.

Blumenthal says that it is too early to tell whether the planned reforms will be enough to reverse this year's decline in student enrolment from abroad. "Whether this is a blip or a trend, we just don't know yet," she says. ■

www.cgsnet.org

T. GUTIERREZ/AP

SOURCE: COUNCIL OF GRADUATE SCHOOLS

Congress may force drug firms to reveal clinical trial data

Erika Check, Washington

Congressional investigators have threatened to enact laws that will make the disclosure of data from clinical trials compulsory.

At a House committee hearing on 9 September, lawmakers said that they were concerned about the way in which drug companies and the Food and Drug Administration (FDA) have handled data about the risks and benefits of the use of antidepressants in children.

Most of the data about potential risks have not been published in journals. The issue caused a furore in the summer, after the FDA demanded that drug companies provide these data and found that children who received some types of antidepressants in clinical trials were more likely to exhibit suicidal tendencies.

Representatives of seven drug companies told investigators at the hearing that they had not deliberately withheld the data. Acknowledging problems with credibility, the companies said they were taking steps to disclose more data. For example, they are participating in a voluntary clinical-trial database organized by the Pharmaceutical Research and Manufacturers Association trade group.

But some lawmakers said that this was not good enough. "I don't accept the idea of a voluntary registry," said Congressman Henry Waxman (Democrat, California) at the hearing. He asked Congress to consider requiring all companies to register clinical trials at their outset in a public database. An international group of medical journal editors announced on 8 September that it would make registration a condition of publication.

Lawmakers also criticized the FDA for stalling congressional investigations. Janet Woodcock, an FDA official, suggested that the law sometimes prevents the FDA from releasing certain information about clinical trials. But Greg Walden (Republican, Oregon) replied: "If the law is preventing you from acting, you need to tell us that, and if it's your own rules that prevent you, you need to fix them."

Concerns about the amount of information released by the FDA were echoed at a meeting of an advisory committee four days after the congressional hearing. "The FDA needs to regain respect," said Maryellen Winter, a mother whose daughter killed herself after taking the antidepressant paroxetine. ■



California poll: voters are split over a plan to issue \$3 billion in bonds to fund stem-cell research.

Critics slate ethical leeway in California stem-cell proposal

Jonathan Knight, San Francisco

Opponents of California's \$3-billion plan to fund embryonic stem-cell research say that the proposal would give researchers carte blanche to rewrite well-established ethical guidelines to suit their needs.

They say the research institute planned under the initiative will be exempt from legislative supervision and, if established, will be able to make its own rules about conflicts of interest and informed consent.

Proponents are reacting angrily to the charges, saying that the proposal provides the highest possible level of accountability and will serve as a model of how science can be funded at the state level.

The California Stem Cell Research and Cures Initiative will appear as Proposition 71 on the ballot in the elections due on 2 November. If passed, it will authorize a bond issue of nearly \$3 billion over ten years to fund embryonic stem-cell research and infrastructure. It will create the California Institute for Regenerative Medicine to distribute the funds.

But public opinion is sharply divided on the proposition, with 45% in favour and 42% opposed. A California columnist has branded it an "audacious raid on the treasury" and details of the initiative are under increased scrutiny in the run up to the election.

The institute will be run by a 29-member Independent Citizens Oversight Committee. Although most committee members will be selected by high-ranking elected officials, including the governor and lieutenant governor, the legislature will have no power to intervene in the institute's affairs or to change the provisions of the initiative.

This could create problems because public money is involved, warns Daniel Sarewitz, an expert on science policy at Arizona State University in Tempe. "Scientists don't like the

fact that the ugly democratic process has stopped them from doing certain kinds of research," he says. "This is an attempt to go around that process."

Proposition-71 backers say that in this case it is necessary to exclude the legislature to ensure that it will not misappropriate the bond fund to help balance the budget. The committee will be accountable in other ways, supporters of the proposition say, with open meetings and an annual analysis of its finances. "We have the highest possible standards of accountability," says initiative spokeswoman Fiona Hutton.

This does not reassure Diane Beeson, a medical sociologist at California State University, Hayward, who says the committee will have unrestricted freedom to rewrite the rules of informed consent. At first, these will have to be "generally based on" the standards established by the National Institutes of Health (NIH), according to the text of the initiative, but the committee will then be free to modify them as necessary.

Beeson claims that passing Proposition 71 will result in poor women lining up to donate eggs for the development of stem-cell lines, in exchange for the modest payments permitted to cover expenses.

Rex Greene, who is an oncologist and spokesman for the opposition, says that the committee will also have the power to rewrite its own conflict-of-interest guidelines, after initially basing them on NIH rules. "It's going to be a sham of peer review," he says.

Such charges are groundless, counters Irving Weissman, a stem-cell researcher at Stanford University, who helped author the measure. He promises that scientists based outside California will advise the committee on which grants to fund. "We want to avoid even the appearance of a conflict," says Weissman. ■

Plan to cull aquarium tuna dead in the water

Rex Dalton, Monterey

Scientists at the Monterey Bay Aquarium in California, one of the world's best-known marine conservation facilities, are struggling to get rid of two huge tuna from the aquarium's main display tank.

Aquarium staff say they tried this summer to kill the healthy, 150-kilogram bluefin tuna (*Thunnus orientalis*), which rule the 4-million-litre Outer Bay tank. Their attempt to dispatch the fish using anaesthetizing darts has riled conservation experts, who say it may breach accreditation rules for US zoos and aquaria.

Marine-husbandry staff at the aquarium say they wanted to remove the fish because they had become too aggressive, particularly towards hammerhead sharks (*Sphyrna lewini*), which have been added in recent years to make the facility's marine environment more realistic.

But conservation biologists elsewhere have questioned the attempt at lethal elimination, because bluefin tuna are an important global conservation symbol and are overfished in the wild. "It strikes me as bizarre," says George Amato, director of the science resource centre at the Wildlife Conservation Society in New York, which has its own aquarium. "That would never happen here."

Officials at Monterey Bay say the attempt to have a veterinarian kill the bluefin was approved by its animal-welfare committee. Darting the bluefin was seen as the safest option for staff, and other species in the tank.



Big problem: Monterey Bay Aquarium staff say their giant bluefin tuna have become aggressive.

The move comes as the aquarium is preparing to introduce a great white shark (*Carcharodon carcharias*) into the tank. The aquarium has been keeping the shark, a juvenile, under observation in another tank. No white sharks have ever been kept successfully in captivity, scientists say.

Randy Hamilton, the vice-president of husbandry at Monterey Bay, insists that the bluefin are not being removed to make way for the new shark. But he accepts that having them out of the way would cut the likelihood of conflict. The shark could be added to the Outer Bay tank as soon as this week — or not at all, if the team feels that the dynamics in the tank are too complex.

Now in its twentieth year, the aquarium is in the midst of a long-term planning process regarding its exhibits. The board will meet on 17 September to consider a plan for the aquarium, which is run by marine

biologist Julie Packard, whose father, computer mogul David Packard, helped to set it up through a charitable foundation.

"Our world is a blend of art and science," says Hamilton. "We would like to increase the visitor experience."

Whatever the outcome, the mix of species for the Outer Bay tank has prompted tensions among scientists at the aquarium and at Stanford University's adjacent Hopkins Marine Station. The two facilities jointly support the Tuna Research and Conservation Center, which has pioneered the use of electronic tags to document where tuna and other species roam.

Stanford's Barbara Block, who directs the tuna research centre, is known to be concerned about the aquarium's handling of the bluefin. But she declined to comment on the aquarium's effort to eliminate the fish, saying only that trying to remove large bluefin tuna "is a hard problem".

Conservation authorities say that having plans for removing animals from research or exhibit facilities is required for accreditation by the Maryland-based American Zoo and Aquarium Association (AZA), which has accredited the Monterey Bay Aquarium.

John Hewitt, director of animal husbandry at the Audubon Aquarium of the Americas in New Orleans, who has served on the AZA's accrediting commission, says that, in general, killing an animal to dispose of it "is not an acceptable practice".

Hamilton says he is now working on a new plan to deal with the bluefin. ■

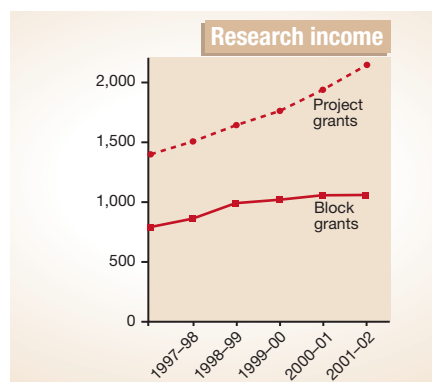
Rule change set to cost Britain Framework cash

Jim Giles, London

British universities say they could be forced to drop out of the European Union's Framework research programme under rules being proposed by the UK government.

The universities say that the government's efforts to change the way it provides their financial support could make it unproductive for academics to seek Framework grants. At the moment, most UK research grants cover only about half a project's true costs, and the main source of the remaining money is block government grants to research institutions. But the block grants have not been keeping pace with the growth in research income (see graph).

So the government announced in July, as part of a ten-year strategy for science, that it would make researchers apply for grants to cover the entire cost of their projects. But



although UK research councils are getting extra funds to help them cope with the switch, the European Commission is sticking to its current arrangements.

Some UK universities might even stop researchers applying for Framework funds because the resulting grants would be unlikely to cover the total project costs,

claims Universities UK, which represents most of Britain's institutes of higher education. "This will bite hard on some departments," says Simon Jones, deputy director for research services at the University of Cambridge.

A drop in applications could hit research hard. Britain currently receives some €700 million (US\$860 million) in Framework funds a year and wins a quarter of all the money allotted to universities by the programme.

Universities want the government to create a new funding stream to help make up the shortfall, which would cost about €100 million a year. Last month, they sent a paper outlining their ideas to the government.

"We hear what is being said," says Tino Hernandez, a spokesman for the government's Office of Science and Technology. "But at the end of the day it is up to universities to make decisions about the research they conduct." ■

Repeal of embryo law urged after child's cure

Federica Castellani

On 6 September, Italy's health minister hailed the successful treatment of a sick child with 'adult' stem cells from newborn siblings. He was hoping to strengthen the government's position that embryonic manipulation is not required for medical progress.

But the minister's declaration rebounded on him the next day, when it emerged that the treatment was only made possible by *in vitro* fertilization of the child's mother in Turkey — using an embryo selection technique that the Italian government outlawed in February.

Now the health minister, Girolamo Sirchia, is facing calls for his resignation, and campaigners are hoping to overturn the law against embryo selection in a referendum, which could be held next summer.

The five-year-old child was treated for thalassaemia, a hereditary blood disease that causes life-threatening anaemia. Doctors cured him by using adult stem cells derived from the umbilical cord blood of twin siblings, borne by his mother.

The parents originally wanted to use cord blood from their next, healthy child, but the blood proved incompatible with their ill son's immune system. So they underwent *in vitro* fertilization in Istanbul, where 12 embryos were created and tested for the presence of the thalassaemia gene and immunological compatibility. Three healthy and compatible embryos were selected for implantation, and twins were born in April.



Girolamo Sirchia: the Italian health minister's gaffe has provoked calls for his resignation.

The boy was treated in Pavia in August. During a visit to the Milan clinic where the cord stem cells were manipulated before implantation, Sirchia told journalists: "This is a historical result, which awakens hopes."

But Italy's law on assisted reproduction bans the testing of embryos for genetic disease, as well as restricting the number of embryos that may be generated for *in vitro* fertilization to three. Now opponents of the law are taking this chance to push for its repeal.

A few days after the law was passed, the minority Radical party launched a petition

for a referendum on whether it should be revoked. The campaign has struggled to collect the 500,000 signatures that would be required before the end of September to oblige the government to hold such a vote next summer. "But this case has given a real boost to the campaign," says Cinzia Caporale, a bioethicist from Rome who opposes the law. "It probably guarantees that the right number of signatures will now be collected."

Sirchia says the campaigners are distorting events for political ends. "The cure of the child and the *in vitro* selection of healthy embryos are not correlated," he says. But scientists point out that, without embryonic selection, the chance of producing an immunologically compatible child who did not have the disease gene would have been less than one in five. Several parliamentarians have called for Sirchia's resignation.

Amid fears that a referendum could threaten the law, a wide spectrum of politicians are trying to negotiate a compromise. Two members of Forza Italia, the party of Prime Minister Silvio Berlusconi, have called for a parliamentary debate about possible amendments on 22 September.

But Francesco D'Agostino, a philosopher at Tor Vergata University in Rome and head of the National Bioethics Committee, which helped formulate the law, is wary. "This specific case is not enough to justify a general law," he says. "To create embryos and destroy those not useful to us is a eugenic practice and we need to think carefully about whether we want eugenics in our law." ■

NASA probes Genesis wreck in bid to salvage data

Nicola Jones, London

The NASA solar-wind probe that crash-landed in the Utah desert last week was cunningly designed to deal with landing problems, mission planners say. But the jury is out on how much data can be saved.

"The prospects for our highest priority objectives are good, although we probably won't be able to do everything we wanted," says Don Burnett, principal investigator on the Genesis project.

The US\$260-million Genesis probe carried delicate wafers of gold, diamond, sapphire and silicon, designed to catch particles from the solar wind. The four plates of these wafers, which were exposed to different kinds of solar activity, were made of slightly different thicknesses. This was intended to help researchers reassemble the plates if a bumpy landing smashed them.

NASA arranged for highly trained helicopter pilots to snag the returning

capsule after parachutes had slowed its return to Earth (see *Nature* 429, 340–342; 2004). Mission planners were fairly confident that damaged plates could be reconstructed if the parachutes worked but the pilots failed, leaving the probe to hit the desert at about 10 kilometres per hour.

Unfortunately, the parachutes themselves failed, and the probe smashed into the ground on 8 September at more than 200 kilometres per hour. "We have a mangled mass of a spacecraft," says David Lindstrom, a Genesis programme scientist based at NASA headquarters in Washington.

Some of the detectors, including a concentrator designed to collect the densest sample of solar wind, seem intact. The impact smashed many of the wafers, however, and also ruptured the container, exposing some, if not all, of the detectors to Earth's air and soil — a much bigger problem for mission scientists.

But they remain optimistic, says Lindstrom. The atoms collected from the solar wind slammed into the detectors at hundreds of kilometres per second, so they should be buried 100–150 nanometres beneath the plate surfaces. It could therefore be possible to distinguish this tiny sample from contamination on the surface.

The mission debris was dug out of the desert and taken to a cleanroom at the Jet Propulsion Laboratory in Pasadena. The Genesis Mishap Investigation Board will try to determine what caused the failure. One possibility is a battery that should have sparked a small explosion to release the parachutes. Another NASA mission, Stardust, is relying on a similar parachute system for its return to Utah in 2006. ■

For more news and analysis go to
news@nature.com
www.nature.com/news



Slippery slope: climate change is driving ski resorts to head for higher altitudes in search of snow.

Ecologists mount protest over lofty plans for Alpine ski runs

Quirin Schiermeier, Munich

As summer comes to a close and millions of skiers eagerly prepare for winter, a heated argument has broken out over plans to build new ski resorts high in the Alps.

The disagreement has been sparked by an initiative to open the second largest glacier in the eastern Alps — the pristine Gepatsch glacier in Tyrol, Austria — to skiers.

Climate change is causing snow lines to creep farther up many mountains, leaving some small ski resorts with access to snow for only a very short time. In response, many ski centres are looking to expand into glacial areas at higher altitudes.

But ski-lift stations, restaurants and piste vehicles produce tonnes of harmful waste, grease, lubricant oils and salts, says Peter Hasslacher, a development planner and conservation expert with the Austrian Alpine Club in Innsbruck. When placed on glaciers, such facilities soon pollute the surrounding soil and water.

Nevertheless, a recent change in Tyrol's environmental law permits the construction of ski lifts in high regions — even in areas that form part of a European network of protected natural landscapes.

Environmentalists have expressed horror at the Gepatsch expansion plan, which they say could ruin one of Europe's last wildernesses. If it goes through, they say, it will open the floodgates for similar projects in other parts of the Alps. "Driving skiers higher and higher is the worst strategy if you want ecologically sound winter tourism," Hasslacher says.

But the tourist industry is facing a difficult future. According to a recent study by the

United Nations Environment Programme, about half of Switzerland's skiing resorts are likely to suffer from a lack of snow in the near future. Within a few decades, the losses to Swiss tourism could reach US\$1.6 billion per year, says the UN report *Climate Change and Winter Sports: Environmental and Economic Threats*.

Leisure companies, such as the French market leader Compagnie des Alpes, are already focusing investment on high Alpine regions, says Michel Revaz of the International Commission for the Protection of the Alps (CIPRA), a non-governmental conservation group based in Liechtenstein. Huge skiing areas, such as those around Zermatt and Saas Fee in Switzerland, are promising a good return on capital, he says.

Environmentalists agree that Alpine communities need to find a way to address their potential losses from climate change. But they would rather they market forms of tourism that have less of an effect on wilderness areas — such as developing conference facilities, or providing support for hiking or sledding in patchy areas of snow at lower altitudes.

"High-altitude skiing is a ruinous waste of energy and natural resources," says Andreas Götz, CIPRA's director. "Clinging to it shows a lack of imagination, and sets a bad example for emerging skiing tourism in places such as China and Russia."

The Alpine Convention, an agreement signed by Austria, Slovenia, Italy, Switzerland, Liechtenstein, Germany and France, does not expressly disallow new high-altitude ski areas. But ministers of the seven countries bordering on the Alps will discuss possible changes to these regulations at a meeting in Bavaria in November. ■

Burnt books get cold shoulder in restoration effort

Karoline Schwarzbach, Munich

Life in the freezer is about to become a reality for thousands of ancient books in eastern Germany.

The books form part of the Duchess Anna Amalia Library in Weimar, one of Europe's most valuable collections, which was badly damaged by fire on 2 September.

Some 30,000 of the million volumes in the library's collection were destroyed in the blaze. And at least 40,000 burnt and water-logged books and manuscripts now need to be restored.

The damaged books are being taken to the Centre for Book Conservation in Leipzig, where they will be stored below -22°C to keep mould and bacteria at bay. They will then be freeze-dried before restoration begins. The federal government has pledged €4 million (US\$5 million) to kick-start the effort, and private donations are already coming in to help complete it.

"Fortunately, the books were packed



Out cold: burnt books from the Anna Amalia Library will be frozen for safe-keeping.

very close together, so the main damage has been to the edges of the pages and on the book covers," says Robert Fuchs, a conservator at the University of Applied Sciences in Cologne.

"The fire was a tragedy for the world of German literature," says Martin Strebel, a restorer who works for the Abbey Library in St Gallen, Switzerland. But he adds that many tools are available to repair the damaged books.

Burnt page edges can be replaced with new paper, for example. And if whole pages look set to crumble, the two sides can be separated and stabilized by a sheet of paper placed in the middle. But only a few experts can carry out this procedure, which is expensive and time-consuming.

Weimar was home to some of Germany's most famous writers, including Nietzsche and Goethe. The library, created in 1761, is listed as a World Heritage Site. ■

Astronomers chase first place in planet snapshot contest

London Two groups of astronomers who think they have taken the first photos of a planet outside our Solar System are now racing to collect data to confirm their finds.

Scientists at the European Southern Observatory reported last week that they had spotted an object orbiting a small, faint star about 230 light years away (G. Chauvin *et al. Astron. Astrophys.* in the press), using a telescope in Chile. Their claim comes just months after reports of a sighting of another possible planet, about 100 light years away, made by a research team at Pennsylvania State University using the Hubble Space Telescope. Their work has been submitted to a journal for review.

Neither group can yet provide definitive proof that the smudgy dots in their snapshots are indeed planets. Both are anxiously waiting for telescope time to make more detailed orbital observations. "The race is on," says Steinn Sigurdsson, one of the Pennsylvania State astronomers.

Although researchers have identified more than 120 possible extrasolar planets by watching for wobbles in the orbits of distant stars, these objects have not been

Indians wade in to stem Chinese flood

New Delhi China has at last decided to let Indian scientists visit a lake on its border with India that is threatening to burst and flood the Indian state of Himachal Pradesh (see *Nature* **430**, 818; 2004). The artificial lake (see right), which holds 60 million cubic metres of water, formed when landslides blocked the course of the Pare Chu river in the Himalayas.

Indian officials have evacuated tens of thousands of people from the danger zone, and made repeated requests to send scientists and engineers to the site. Those requests were finally granted on 10 September. An Indian team of geologists and hydrologists, led by the home secretary, plans to leave soon for China to discuss measures to prevent the dam bursting in time for the next rainy season in June 2005.



imaged directly. Photos should help astronomers pin down the identities of these mystery objects.

Primate lab charged with cruelty to chimpanzees

San Diego A New Mexico primate facility has been charged with animal cruelty by the local district attorney.

The Alamogordo Primate Facility (APF) houses a few hundred retired research animals, such as chimps that were used in HIV experiments. Animal-rights groups have

long had concerns about the facility, which was previously under the management of the Coulston Foundation. Now, In Defense of Animals has successfully lobbied district attorney Scot Key, and charges have been brought against head veterinarian Rick Lee and Charles River Laboratories, which runs the facility under contract from the National Institutes of Health.

The charges claim that three chimps were left overnight without proper care, leaving one to apparently choke to death on its own vomit and another to die from its wounds.

Publisher, Physical Sciences, NPG

Nature Publishing Group is seeking a publisher to develop and expand the group's programme of publications in the physical sciences. This programme includes *Nature Materials*, which less than two years after launch has achieved the world's highest impact factor for a physical science journal at 10.8, *Nature Chemical Biology*, due to be launched in Boston in June 2005 with an additional editorial office in San Francisco, and *Nature Physics* to be launched in London in October 2005.

The publisher, who will be based in either Boston or London with frequent travel between the two locations, will be responsible for these publications and for planning and launching new titles both online and in print in other areas of the physical sciences including chemistry.

This is a senior position with an attractive salary package. It requires sound experience in science publishing, strong commercial drive, and knowledge and interest in the physical sciences and science in general.

This position will report into David Swinbanks, Publishing Director, Nature Publishing Group,

To apply applicants should provide, along with a CV and letter of application, a short document of 1000 to 1500 words outlining their strategy to develop a series of titles in the physical sciences in the next three to five years' building on those already planned or published. The document should cover both online and print publishing opportunities.

Please send applications, quoting reference number NPG/LON/152, to Sarah Pilkington, Personnel Administrator, Macmillan Publishers at londonrecruitment@macmillan.co.uk

Closing Date: 24 September 2004

nature publishing group 

A statement from Charles River Laboratories says that “in all three cases, APF veterinarians provided the immediate and appropriate medical attention necessary to the animals, all of whom had underlying health issues because of the diseases to which they had been exposed”.

House vote singles out mental-health projects

Washington The US House of Representatives voted to take cash away from two studies of undergraduate behaviour, during a 9 September debate on the National Institutes of Health (NIH) budget.

An amendment was passed to strip funding from two National Institute of Mental Health projects — one examining how college students decorate their dormitory rooms, the other looking at the relationship between college students' goals and their level of happiness.

The amendment serves as a protest in words only — money won't actually be taken from either project, as the funding has already been handed out. There have been previous attempts to take money away from specific NIH projects: last year a measure to take the cash from a number of grants, most of which had to do with sex, was narrowly defeated in the House.

NASA's mission to the Moon gets bigger price tag

Washington President Bush's plan to send people back to the Moon and on to Mars could cost an astronomical \$127 billion.

The Congressional Budget Office (CBO), which provides non-partisan accounting analyses to US law-makers, estimates that the programme will cost 33% more than NASA has predicted. And that's just to get to the Moon.

NASA's estimate was \$66 billion for the human exploration programme up to the



Moon over Washington: rising cost estimates cast a shadow over NASA's space programme.

next lunar landing in 2020. A further \$29 billion would go on robotic support missions.

The CBO came to its estimate by looking at the budget record of 72 previous NASA missions. But such numbers are inevitably hard to pin down for a mission that everyone concedes is still fairly loosely defined. NASA hopes to unveil details of the proposed Crew Exploration Vehicle, the first hardware element of the programme, later this month.

Society admits paper was published in error

Washington A peer-reviewed journal that published a paper advocating creationist ideas has said that it erred, and vowed not to print similar ideas again.

The little-known *Proceedings of the Biological Society of Washington* found itself the centre of attention last week, as evolutionary theorists rounded on its decision to publish a paper on intelligent design, a theory that rejects evolution (see *Nature* **431**, 114; 2004).

A statement issued on 7 September said that the article was published without the knowledge of the society's council, who had now deemed it inappropriate. The paper was accepted under the editorship of Richard Sternberg, who has since left the journal.

Islet Cell Transplant, Prelude to the Future

In Honor of Paul E. Lacy in his 80th Year

Co-Chairmen

Allen Spiegel, M.D.
Director, National Institute of Diabetes & Digestive & Kidney Diseases, National Institutes of Health

Walter Cohen, D.D.S.
Chairman of the Board, National Disease Research Interchange
Chancellor Emeritus, Medical College of Pennsylvania, Hahnemann University

Keynote Speaker

James Shapiro, M.D., Ph.D.
Director, Clinical Islet Transplant Program, University of Alberta, Edmonton

Historic Perspective/ Islet Cell Transplant

Paul E. Lacy, M.D., Ph.D.
Chairman Emeritus, Department of Pathology, University of Washington

Join our roundtable discussions with these scientists:

"Obstacles and Challenges to Islet Cell Transplant"

Co-Chair: Paul Robertson, M.D.
Chief Executive Officer, Pacific Northwest Research Institute
Co-Chair: Camillo Ricordi, M.D.
Scientific Director and Chief Academic Officer,
Diabetes Research Institute, University of
Miami School of Medicine

"Alternative Paths to the New Therapy, the Face of the Future"

Chair: Ali Najj, M.D., Ph.D.
Professor of Surgery, University of Pennsylvania

Presentations:

Paul E. Lacy Medal for "Outstanding
Scientific Achievement" to
James Shapiro, M.D., Ph.D.
Presenter: Lee Ducat
President, NDRI

December 1 - 2, 2004

Rittenhouse Hotel
Philadelphia, PA

For conference information and to make a reservation, visit the NDRI web site at www.ndriresource.org. Reserve now, limited seating available.

NDRI
National Disease Research Interchange
THE NATIONAL RESOURCE CENTER



Head to head

The party conventions are over, and the candidates have been anointed. Now it's a straight race to the tape between President George W. Bush and his challenger John Kerry. *Nature* asked them where they stand on science.

In their frantic search for votes, George W. Bush and John Kerry are leaving no stone unturned. They've even been sparring over science, with Kerry raising the issue of embryonic stem-cell research at the Democrats' convention in July, and Bush accusing Kerry of flip-flopping on the site of a nuclear waste dump in Nevada soon after.

Sections of the scientific community have pitched in, with Kerry's campaign attracting vigorous support from a group backed by 48 Nobel laureates. The fact that most of this activity is on the Kerry side says a lot about recent political trends in the United States. This is a deeply divided country, and scientists and the university campuses that many of them inhabit have become distinctly unfriendly territory for the Republicans.

But scientists pride themselves on objectively assessing evidence. So, for the first time in *Nature's* history, we have given the candidates the chance to address researchers directly. From about 50 questions posed by our editorial staff, we selected 15 and asked the campaigns to respond in 1,500 words, distributed as they saw fit. Bush's answers were some 30% over length, and have been edited; Kerry kept to the limit and his responses are presented in full.

We hope that the result will give an inkling of what the candidates stand for. In some areas, such as the broad balance of science funding, there is not much to choose between them; in others, such as global warming, their respective stances could hardly be further apart.

These stances reach beyond domestic US issues. As the Hungarian-born financier George Soros observed two years ago: "In modern global capitalism, only Americans vote, Brazilians do not." What he meant was that when it comes to economic, military and other decisions, the US administration's actions are likely to have as much impact on your country as those of your own government.

That's why, this November, from the remotest province of China to the bustling capitals of Europe, so many eyes will be on America's votes. Let's hope they count them up right.

Colin Macilwain, news editor

See news@nature.com for more election coverage ▶ www.nature.com/news



Bush vs Kerry



initiative to support nonproliferation, disarmament, counterterrorism and nuclear safety projects in the former Soviet Union. For nuclear weapons, the first step is to prevent access to fissile materials. We are making good progress in this area through efforts such as the Global Threat Reduction Initiative and our material security efforts in Russia.

KERRY: It is not a problem we will be able to solve alone. It is going to require American leadership that forges an international consensus on how to deal with these weapons and the often dual-use technology that underpins them. I will work closely with the scientific community to develop responsible oversight for biomedical research to make sure that deadly pathogens are only in the hands of those with legitimate research needs. Together we will find ways to reduce the possibility that scientific knowledge and capabilities will be misapplied to do harm.

4 Do you support research into new nuclear-weapon designs in the United States? If not, how do you see the future role of the three nuclear-weapons labs?

BUSH: Our national laboratories are doing great work to deal with the threats of the twenty-first century. These laboratories are also a tremendous asset in our efforts to improve homeland security, are the source of unparalleled technological progress, and are helping America win the War on Terror. The Nuclear Posture Review released by my administration in January 2002 noted that the nation's nuclear infrastructure had atrophied since the end of the cold war and that the evolving security environment requires a flexible and responsive weapons-complex infrastructure. To that end, my fiscal-year 2005 budget reflects an increase over 2004 in weapons activities.



1 Is there a danger that increased controls on travel by scientists into the United States, introduced in response to homeland security concerns, will isolate US science and endanger US scientific leadership? If so, what can be done to keep US science open to the world?

BUSH: My administration values the contributions that foreign scientists and students make to our nation's scientific enterprise, while recognizing the importance of safeguarding our security. We will continue to welcome international students and scientists while implementing balanced measures to end abuses of the student visa system.

We have already achieved several notable successes in reducing delays now being experienced by some visa seekers. We have increased security while speeding up the clearance process — about 1,000 backlogged applications have already been cleared out.

KERRY: We can balance science and security. In the wake of 9/11, America took important steps to improve security for visa applicants to the United States. However, we can improve our visa system to process visa applications for legitimate scientists and students more quickly while still screening individuals who pose a genuine security risk. With more resources and better procedures, we do not need to face a trade-off between scientific exchange and national security.

2 Recent months have seen various charges of political bias against scientific panels that advise the US government at different levels. What would you do to

ensure that your administration receives genuinely impartial scientific advice?

BUSH: My administration has a strong commitment to the highest scientific standards in decision-making. On issues ranging from climate change to nanotechnology, I have sought out the best scientific minds — inside and outside the government — for policy input and advice, especially the independent National Academies. My commitment to sound, independent scientific advice is unwavering. And my senior science adviser in the White House, John Marburger, happens to be a Democrat.

KERRY: My administration would never utilize biased advice as a foundation for public policy. As president, I will serve on behalf of the public interest. In order to best serve the public, effective decisions must be made with the input of genuine impartial expert counsel.

3 What is the long-term solution to the gradual dissemination of knowledge about weapons of mass destruction — especially bioweapons?

BUSH: Stopping the gradual dissemination of knowledge is impractical if not impossible. The key to stopping the proliferation of weapons of mass destruction is preventing those seeking these weapons from gaining access to their most significant and technically challenging components. The redirection of former weapons scientists to productive civilian employment is a key priority.

My administration has launched the G8 Global Partnership — a \$20-billion

KERRY: I would end the pursuit of a new generation of nuclear weapons. Our national laboratories play a critical role in maintaining our existing stockpiles and assuring that our existing nuclear weapons are safe, secure and reliable. They also play and should continue to have an important role in preventing the spread of weapons of mass destruction and in advancing science for our nation's security.

5 Some physicists have questioned the capability of missile defence systems being deployed in the United States. Would you increase or decrease spending on missile defence, and would you subject claims made on its behalf to independent scientific review?

BUSH: Early in my administration, I called for the examination of the full range of available technologies and basing modes for missile defences that could protect the United States, our deployed forces, and our friends and allies. Our policy is to develop and deploy, at the earliest possible date, ballistic missile defences drawing on the best technologies available.

Later this year, the first components of America's missile defence system will become operational. This will fulfil a pledge I made to the American people more than four years ago. We will develop and deploy the technologies necessary to protect our people.

KERRY: I am not for rapid deployment of missile defence. We should not waste money on deployment at this point. I favour additional research, development and testing. As to the issue of independent scientific review, we have to be careful because of the classified nature of much of the work in question. At the very minimum, we must work hard to restore the credibility of the internal review process. The truth is the Bush administration has shredded its own credibility on this, particularly in its rush to deploy missile defence. We need to restore the credibility of our own review process and we need to subject systems to realistic, operational testing to make sure that they really work.

6 Should the United States participate fully in the construction of ITER, the proposed fusion research facility, and

what steps would you take to help such international scientific projects to succeed?

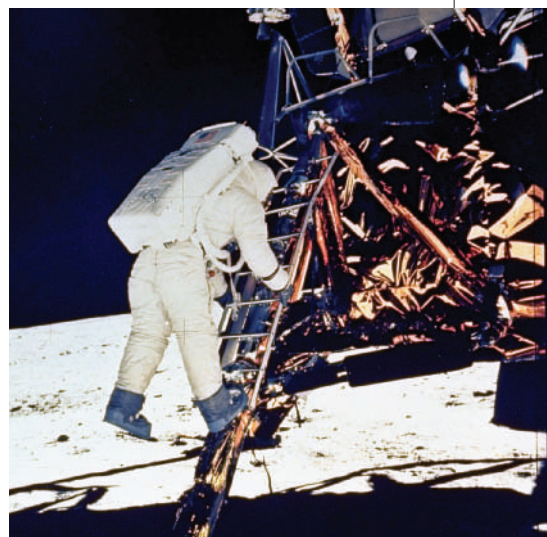
BUSH: I committed the United States to join ITER early in 2003. ITER is a critically important experiment to test the feasibility of nuclear fusion as a source of electricity and hydrogen. Along with several other nations, the United States is playing a critical role in launching ITER. In fact, ITER is the Department of Energy's top facilities priority.

KERRY: My energy plan will tap America's initiative and ingenuity to strengthen our national security, grow our economy and protect our environment. With regard to ITER, John Edwards and I support a strategically balanced United States fusion programme that includes participation in ITER to supplement a strong domestic fusion science and technology portfolio. As president, my first priority internationally on this and other energy issues will be to engage other nations to find areas of cooperation and common ground.

7 Do you think the United States should send astronauts to the Moon or Mars in the next 10 to 15 years? If so, why send humans instead of robots? If not, what is the purpose of the space shuttle and space station?

BUSH: In January, I announced my vision for the future of America's space exploration programme. As we complete our work on the International Space Station, we are developing a new manned exploration vehicle to explore beyond our orbit. This vehicle will be tested by 2008 and will conduct its first manned mission no later than 2014. America will return to the Moon as early as 2015 and no later than 2020, and use it as a foundation for human missions beyond the Moon. We will begin with robotic missions, and manned missions will follow. An extended human presence on the Moon could reduce the costs of further exploration.

KERRY: Today, thanks to decades of public investment in space exploration activities, a rotating international team of astronauts is living and working in space on the International Space Station, a dozen Americans have walked on the Moon, we have rovers exploring the surface of Mars and an armada of spacecraft continues to



explore our Solar System. NASA is an invaluable asset to the American people and must receive adequate resources to continue its important mission of exploration.

However, there is little to be gained from a space initiative that throws out lofty goals, but fails to support those goals with realistic funding. John Edwards and I are committed to increasing funding for NASA and space exploration because it not only makes critical contributions to our economy, it also expands our understanding of the world we live in.

8 Some researchers have expressed concern over what they see as a growing disparity between funding for biomedical research and other fields, including the physical and environmental sciences. Do you agree that this a problem and, if so, what would you do about it?

BUSH: My administration is committed to funding basic research and has listened to concerns from the scientific community and lawmakers to ensure that there is a federal priority on funding for physical sciences as well as life sciences. My budgets have sent a strong signal that we are addressing the concerns.

KERRY: John Edwards and I would increase federal funding at both the National Institutes of Health and the National Science Foundation (NSF). To ensure we remain strong in the sciences and engineering, I would specifically increase NSF funding for the physical and environmental sciences, and double the NSF graduate scholarships for mathematics and science.

We must not short-change our national investment in future medical and technological breakthroughs. It will be scientific discoveries that will drive our

Bush vs Kerry

future economy — just as the discoveries of electricity, the combustion engine and the Internet drove our economy in the past.

9 Many environmental problems can be attributed to high levels of consumption in developed nations such as the United States. Can science and technology allow everyone on the planet to reach these levels of consumption? Or do Americans need to change their lifestyles and consume less?

BUSH: America in a very real sense has changed, not by consuming less, but by consuming and producing smarter. We have proven that economic growth makes possible the environmental progress our country has achieved and will continue to achieve in the future.

Under my leadership, America has entered productive international partnerships to assist developing countries in building more modern energy systems. Given the enormous gains of the past century, I do not and would not underestimate the enormous potential of science and technology to continue to make possible improved living standards for people all over the world.

KERRY: John Edwards and I believe that we can protect our environment while strengthening our economy. Time and time again, America has met environmental challenges through ingenuity and technological innovation. But it takes strong leadership to put the interests of protecting public health and the environment ahead of the interests of polluters, and as president I will reverse the four years of environmental neglect by the Bush administration. I have been a leader in the fight to strengthen our

economy and protect our environment, fighting to clean up toxic waste sites and to keep our air and water clean.

10 Does the Endangered Species Act need to be amended in order to operate more effectively? If so, how would you amend it?

BUSH: We need to modernize the act so that it provides the greatest benefits to those species most in need. For example, productive reforms could include habitat conservation plans, conservation banking, voluntary agreements with landowners, and partnerships with states, tribes and nongovernmental organizations. These programmes could provide far greater conservation benefits while avoiding unnecessary regulatory, economic and social burdens.



KERRY: John Edwards and I support protecting wildlife and the important goals of the Endangered Species Act. We will implement the act in a cooperative manner that extends the benefits of wildlife and habitat protection to public and private lands. With adequate funding and a cooperative approach that works for both wildlife and property owners, we will continue America's strong legacy of protecting wildlife.

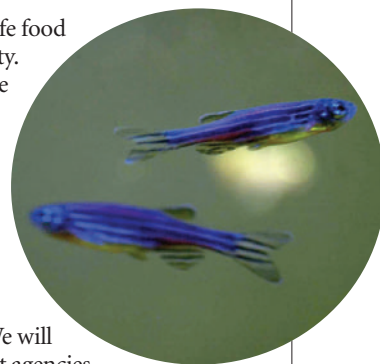
11 Most Americans accept transgenic crops as safe to eat, transgenic salmon are being developed for sale as food, and genetically modified fish that glow in the dark are being sold in pet shops. At what point does genetic modification of plants and animals become problematic to you?

BUSH: Biotechnology plays an extremely important role in reducing environmental impacts of farming and meeting the world's increasing demand for food. But I believe it is important that our regulatory framework keeps pace with science. The agriculture department's Animal and Plant Health Inspection Service has begun developing a wide-ranging environmental impact statement to assess the effectiveness of biotechnology regulations. This will help the federal government better understand risks and benefits.

KERRY: John Edwards and I will work towards the goal of reducing the ecological footprint of agriculture and ensuring

adequate and safe food and sustainability.

We will redouble government efforts to make sure biotechnology is safe for human consumption and safe for the environment. We will give government agencies the power they need to effectively regulate genetically modified food products, both before and after market. And we will work with the international community to effectively address its concerns and improve trade relations.



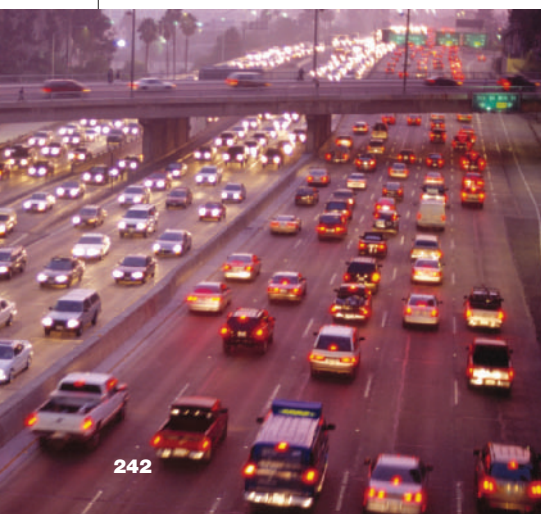
12 Are greenhouse gases generated by the burning of fossil fuels the main cause of global climate change? Is this an important problem for the United States and, if so, what would your administration do to limit emissions of greenhouse gases at home and abroad?

BUSH: Global climate change is a serious long-term issue. In 2001, I asked the National Academy of Sciences to provide the most up-to-date information about the science of climate change. It found that considerable uncertainty remains about the effect of natural fluctuations on climate and the future impacts climate change will have on our natural environment.

My administration is now well along in implementing a comprehensive climate-change strategy to advance the science, expand the use of transformational energy and carbon sequestration technologies, and mitigate the growth of greenhouse-gas emissions in the United States and in partnership with other nations.

I created the new US Climate Change Science Program (CCSP) to refocus the federal government's climate research programmes. The National Academy endorsed the CCSP, noting that it "articulates a guiding vision, is appropriately ambitious, and is broad in scope". I also committed the nation to a goal of reducing American greenhouse-gas intensity by 18% over the next ten years.

KERRY: The scientific evidence is clear that global warming is already happening and rising levels of global warming pollution are making the problem worse. For years in the Senate, I have worked with our allies to fight for a balanced global warming treaty. President Bush rejected the Kyoto Protocol, stubbornly walking away from



the negotiating table altogether. John Edwards and I will take the United States back to the international negotiating table while working at home to take concrete steps to reduce pollution, setting concrete limits to halt and reverse the growth in global warming pollution and tapping the ingenuity of American industry.

13 The Food and Drug Administration (FDA) constantly has to balance the desire for rapid approval of new drugs against the need to ensure their safety. Is the current system getting this balance right? If not, how does it need to change?

BUSH: Today, the FDA sets the world's gold standard for speeding new therapies to patients and ensuring the safety of the drug supply. In 2003, the FDA approved 466 new and generic drugs and biological products, while decreasing the time it took to review and approve most applications. In addition to evaluating new drugs for safety and efficacy, the FDA is now directing monitoring efforts to the 10,000 drugs that are already on the market.

KERRY: As president, I will ensure that the FDA has the resources it needs to approve drugs in a safe and timely manner. In the US Senate, I sponsored and supported legislation that requires drug manufacturers to pay fees to the FDA and allows the agency to hire more reviewers and significantly accelerate drug reviews and approvals. Under the Prescription Drug User Fee Act, new drugs are being approved rapidly by the FDA — and I believe that more should be done to assure the safety of those drugs once they are marketed.

The biggest threat to our success in expanding patient access to medical breakthroughs is the Bush administration's ideological approach to scientific decision-making. When it comes to the safety of our medicines and food supply, the public health is taking second place to special interests and ideological agendas.

John Edwards and I support a return to sound science at the FDA and throughout the federal government.

14 Is mad cow disease, and its possible transmission to people, a significant potential public health threat in the United States? If so, what steps would you take to ensure its containment?

BUSH: My administration is taking aggressive actions to protect American consumers against

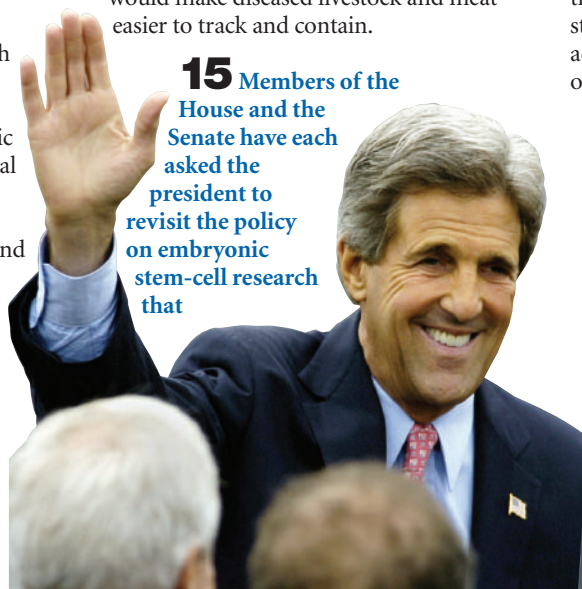
so-called 'mad cow' disease (BSE). Last December, after the discovery of BSE in one cow in Washington state, the Department of Agriculture (USDA) took steps to further assure the safety of our beef, including banning from the human food chain so-called 'downer cattle' and prohibiting specified risk material from animals over 30 months in age. I also called on an international review team to assess the BSE situation and at its recommendation USDA launched an expanded surveillance programme of high-risk cattle populations.

KERRY: Reports of the Bush USDA's mishandling of mad cow disease have underscored the need to protect American consumers. John Edwards and I will improve our food safety and inspection process,



including not only the meat-packing side but also more coverage to ensure that the 1997 feed ban on ground bone-meal is not being violated. We will increase the testing and inspections overall and enhance the BSE surveillance programme so that all suspect animals are analysed. And we will implement an aggressive timeline for establishing a national tracking system that would make diseased livestock and meat easier to track and contain.

15 Members of the House and the Senate have each asked the president to revisit the policy on embryonic stem-cell research that



was announced on 9 August 2001. If elected, would you change this policy and, if so, how?

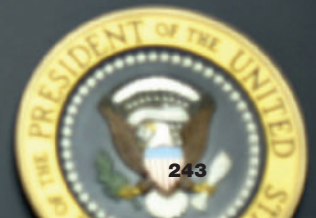
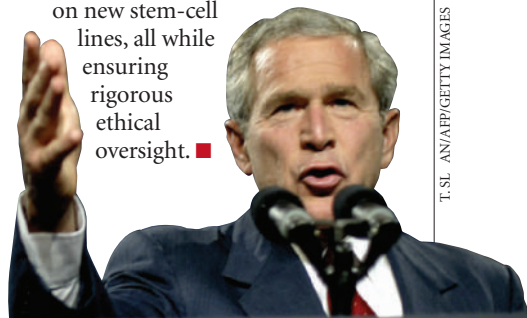
BUSH: I am committed to pursuing stem-cell research without crossing a fundamental moral line, and I am the first president to provide federal funding for human embryonic stem-cell research. However, stem-cell research is in a very early stage and while it may hold great promise we should not overstate the state of the science, or politicize these issues, because it gives false hope to individuals and families suffering through terrible illnesses.

Last year, the federal government invested \$25 million in embryonic stem-cell research and nearly \$191 million in adult stem-cell research. And these efforts are matched with millions more dollars spent in the private sector. My administration is also creating a national embryonic stem-cell bank.

These efforts are providing a boost to research in a very promising new field, while not providing taxpayer funding that would sanction or encourage further destruction of human embryos. My policy makes it possible for federally funded researchers to explore the potential of embryonic stem cells, while respecting the ethical and moral implications associated with this research.

KERRY: Today, millions of children and adults suffer from incurable diseases such as diabetes, Parkinson's, Alzheimer's, heart disease, cancer and spinal-cord injuries. John Edwards and I believe that we must lift the barriers that stand in the way of science and push the boundaries of medical exploration so researchers can find the cures that may exist. I will lift the ideologically driven restrictions on stem-cell research created by the Bush administration by overturning the ban on federal funding of research on new stem-cell

lines, all while ensuring rigorous ethical oversight. ■



Crick and Darwin's shared publication in *Nature*

A humble cockle and the family link between two minds that explored the origins of life.

Sir — Many of the published eulogies to the recently departed Francis Crick, mine included, have compared him to other scientific greats, such as Charles Darwin. I discovered only recently that Charles Darwin's last publication was, in effect, a joint publication with Francis Crick's grandfather.

On 18 February 1882, Walter Drawbridge Crick, an amateur malacologist and professional shoe manufacturer living in Northampton, wrote to Darwin to say that he had found a small freshwater cockle attached to the leg of a water beetle. Darwin was always interested in how freshwater animals, and molluscs in particular, dispersed by hitch-hiking on other animals.

The issue mattered because freshwater invertebrates vary surprisingly little from

one region of the world to another. This could mean either that the diffusion of freshwater shells "took place before the present distribution of land and water", as suggested by John Gwyn Jeffreys in his *British Conchology* (van Voorst, London, 1862–1869) — or that there is frequent dispersal and population mixing, as argued by Darwin.

Crick's grandfather guessed rightly that Darwin would be interested in the water beetle's passenger. Darwin replied with a barrage of questions. Crick sent him the beetle and the shell, both of which survived the journey. Darwin sent the shell to Gwyn Jeffreys for identification, but Gwyn Jeffreys was away from home and the shell was returned by a servant, broken. Crick, who knew his molluscs, had already identified it as *Sphaerium corneum*, which

Darwin knew by its synonym of *Cyclas cornea*. Crick, meanwhile, had returned to the pond where he caught the beetle and found a dead frog with a bivalve of the same species clamped to its foot.

On 6 April 1882, Darwin's note "on the dispersal of freshwater bivalves" appeared in *Nature*. As well as describing Crick's peripatetic cockles, Darwin recalled the extraordinary fact that he had caught a freshwater beetle (of a different genus) while on HMS *Beagle*, 45 miles from land.

Thirteen days later, Darwin died. Crick died in 1903 at the age of 46, fifty years before his grandson co-discovered a cosmopolitan, universal code shared by all living creatures.

Matt Ridley

Blagdon, Seaton Burn,
Newcastle upon Tyne NE13 6DD, UK

It's the science that's a disaster in the movies ...

Sir — Your News Feature "Hollywood or bust" (*Nature* 430, 720–722; 2004) looks at the latest attempts to wed science and Hollywood. As a newspaper science writer, I would urge that any such marriage be quickly annulled.

In 1996 I wrote a non-fiction popular book on the science of tornado and severe storm research. Being a film buff, I was delighted when the book was acquired as the official tie-in book for the Warner Brothers film *Twister*.

Later, to my embarrassment, I saw the film and was horrified by its manglings of meteorological science and terminology. And yet these same film-makers had bragged about "consulting" respected meteorologists.

Even a scientist as influential as Carl Sagan struggled, not entirely successfully, to preserve verisimilitude in the film version of his novel *Contact*. The movie *Contact* omits the novel's most haunting premise, which concerns the value of π . Why? Because (as one of the film's top producers later assured me) π — a concept taught in every American high school — is too difficult for audiences to grasp. What remark better expresses Hollywood's contempt for its audience?

And it's only getting worse. As film budgets soar, a studio's future might depend on the success of a single blockbuster. Thus scriptwriters are increasingly pressured to erase any dialogue that might bore or confuse lowbrow viewers. In a

business as cruelly capitalistic as filmmaking, what alternative do they have?

As the late John Gregory Dunne observed, on the basis of his many Hollywood experiences, no aesthetic or intellectual argument can withstand the movie mogul's favourite counterblast: "It's our money."

If Hollywood calls your laboratory, hang up.

Keay Davidson

San Francisco Chronicle, 901 Mission Street,
San Francisco, California 94103, USA

... yet even flawed films raise interest in research

Sir — We read with great interest your News Feature "Hollywood or bust" (*Nature* 430, 720–722; 2004) on scientists attending a screenwriting class. We use movies to teach non-science students some basic science in a course called Science and Cinema.

Movies are a wonderful medium for sparking the interest of students. We use films such as *Outbreak* to teach students about viruses, *Jurassic Park* to discuss cloning, *Gattaca* for genetic screening and next year *The Day After Tomorrow* will be used to introduce the science of climate change. Those familiar with these films will agree that the cinematic quality is variable and the representation of the science sometimes flawed — but all the films build an exciting story around a scientific centrepiece, and hence are an excellent teaching tool.

As scientists, we understand the concern that movies may misinform the public and that the profession is not accurately represented on the screen. We need to keep these things in perspective, however. Movies usually tell larger-than-life stories and exaggerate characters. Is the work of a lawyer more accurately portrayed in films than that of a scientist? Scientists must join a long list of professions stereotyped by the movie industry.

Furthermore, to assume that most viewers don't understand that movies distort science in the same way that they distort historical events is somewhat patronizing.

In Cinema and Science, the students are taught some basic principles of a discipline; spotting the flaws in the movie is for many participants the most rewarding bit. We find it is often the flaws that inspire non-science students to want to know more about the current scientific research, the future possibilities and the responsibilities that come with them.

Let's use the excitement of movies to increase the understanding of science in the general public, rather than focus on the inevitable errors.

J. Justin Gooding, Katharina Gaus

School of Chemistry and Centre for Vascular Research, The University of New South Wales, Sydney, New South Wales 2052, Australia

correspondence

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. Published contributions are edited.

Economic interests

Do strangers cooperate when they have to work together?

The Company of Strangers: A Natural History of Economic Life

by Paul Seabright

Princeton University Press: 2004. 320 pp.

\$29.95, £19.95

Herbert Gintis

Edward O. Wilson's call for the unification of biology and the social sciences some three decades ago came in for some rough treatment, and the notion of 'sociobiology' is still opposed by some traditionalists. Yet, despite this hostility, the process of integrating social science into natural science seems to be in full swing. Paul Seabright's new book is a welcome and important contribution to this process.

The idea behind sociobiology is that there are many social species, and our understanding of ourselves will be enhanced by analysing the similarities and differences between human and non-human social systems. The title of Seabright's book, *The Company of Strangers*, isolates a unique characteristic of human sociality: although several species have evolved a highly complex and decentralized division of labour, humans are the only species with extensive cooperation among unrelated individuals.

The maturation of sociobiology since Wilson's call to arms has included several key strands of research. One is a broadened concept of sociality which recognizes that, from the emergence of multicellular organisms to the rise of *Homo sapiens*, major evolutionary transitions have required new mechanisms to bring about cooperation among the complex parts of biological entities. It is now routine, for instance, to note that the disciplining of an aberrant cell in an organism, an ovipositing worker in a beehive and a shirking worker in a business enterprise are done in a similar way. A second contribution is gene-culture coevolutionary theory, which treats culture as a form of informational transfer across generations, subject to much the same evolutionary forces as genes. This is important because human sociality has been far more cultural than that of any other species.

The Company of Strangers exemplifies a new breed of economic analysis, seeking answers to fundamental questions wherever they are found and ignoring disciplinary boundaries. A transdisciplinary approach to economic life is nothing new: Adam Smith, for instance, wrote not only *The Wealth of Nations*, but also *The Moral Sentiments*, perhaps the greatest work of psychology before William James. But this tradition was all but buried in the early years of the twentieth century, to be rediscovered only recently.



Seabright provides elementary, but nonetheless richly fascinating, introductions to such standard economic topics as the division of labour, prices, money and commodities. And he addresses such perennial economic problems as unemployment, poverty, environmental destruction and economic instability. The novelty is that he consistently does so from a long-term evolutionary perspective. This is decidedly not a book on economic policy. Such traditionally central questions as capitalism versus socialism, the balance between competition and regulation, and the distribution of wealth and income are mentioned only in passing.

The book's innovation lies in its treatment of the psychological prerequisites of modern economic life. As Seabright notes in the introduction: "Modern society is an opportunistic experiment, founded on a human psychology that had already evolved before human beings ever had to deal with strangers in any systematic way." This psychology has two elements. The better known part is what Seabright calls "rational calculation", by which he means a capacity for logical reasoning, information processing and mastery of technique that far exceeds that of any other animal. Less well known in behavioural science is what Seabright calls "reciprocity", which he defines as "the willingness to repay kindness with kindness and betrayal with revenge, even when this is not what rational calculation would recommend".

It is important to be clear on two terminological issues from the outset. First, by "reciprocity" Seabright means what I and others have referred to in *Nature* as "strong reciprocity". The "strong" adjective is meant

to distinguish the behaviour from the self-interested notion of reciprocity common in the biological literature. Second, Seabright follows a long tradition in economics of considering reciprocity to be non-rational, using the term "rational" to mean "caring only about oneself". There is nothing irrational about such elements of strong reciprocity as returning kindness with kindness and retaliating against someone who has harmed one, however, even when these behaviours involve net material costs.

Seabright's treatment of human society is innovative because biologists and economists alike have long maintained both that humans are selfish when dealing with non-kin, and that their cooperation can be explained by long-term self-interest. Moreover, there is a long tradition, especially on the political left, of criticizing capitalism for promoting greed and selfishness; this is at best a partial truth, as market economies at least tolerate, and probably promote, fair-minded behaviour. Experimental economics, as described here by Seabright, has shown that most people are indeed reciprocal, and that economic and biological models of self-interested cooperation are rarely plausible when they involve groups of more than a few individuals.

Seabright also analyses the dark side of strong reciprocity, which is the tendency to exhibit hostility to "outsiders" in the name of "insider" cooperation. "Cooperation within a group," he observes, "can make the group more lethally aggressive in its dealing with outsiders." The systematic killing of unrelated individuals is so common among humans, he adds, that it "cannot be described as

exceptional, pathological, or disturbed". He concludes that "what Adam Smith famously described as the human propensity to 'truck, barter and exchange' has always coexisted uneasily with a rival temptation to take, bully, and extort."

The Company of Strangers is highly readable and will be accessible to a wide audience. It is, however, weak on detail and eschews formal model building and extended analytical argumentation. As a result, it will serve only as a stepping-stone to the field for those interested in the economy as a dynamically evolving system. ■

Herbert Gintis is at the Santa Fe Institute and is emeritus professor at the University of Massachusetts, 15 Forbes Avenue, Northampton, Massachusetts 01060, USA.

Hidden history

Adolf Butenandt und die Kaiser-Wilhelm-Gesellschaft: Wissenschaft, Industrie und Politik im "Dritten Reich"

edited by Wolfgang Schieder &

Achim Trunk

In German

Wallstein: 2004. 456 pp. €34

Benno Müller-Hill

Adolf Butenandt was the greatest non-Jewish German biochemist of the past century. In 1933, at just 30 years of age, he became a professor at the University of Danzig. At 33 he turned down an offer to become a professor at Harvard and accepted the directorship of the Kaiser Wilhelm Institute of Biochemistry in Berlin-Dahlem, his predecessor, Carl Neuberg, having been sacked by the Nazis on account of his Jewish ancestors. Butenandt went on to receive a Nobel Prize in 1939 for his work on female sex hormones.

Together with Alfred Kühn, Butenandt solved the biochemical problem of the eye colour of mutant *Ephesia* moths, and convinced the chemical industry to finance his research. After the Second World War he became an influential figure in the Max Planck Society, which succeeded the Kaiser Wilhelm Society, serving as its president from 1960 to 1972. A biography of Butenandt by his former student Peter Karlson was published in 1990 but was little more than a hagiography: Butenandt was portrayed as being an immaculate scientist.

But in the records of the DFG, the country's main research funding agency, lurks another story, which came to light in 1983. The human geneticist Otmär von Verschuer told some colleagues in 1946 that his collaborator Josef Mengele had sent him 200 blood samples from the Auschwitz concentration camp, and that Günter Hillmann, who worked with Butenandt, helped with



Top secret: Adolf Butenandt tried to destroy evidence of his research.

the analysis. Another of Butenandt's colleagues, Gerhard Ruhenstroth-Bauer, had meanwhile been helping Hans Nachtsheim to test children for epilepsy in a low-pressure chamber belonging to the Luftwaffe. During Butenandt's lifetime, the Max Planck Society turned a blind eye to these facts. But two years after Butenandt's death in 1995, Hubert Markl, the society's president at the time, set up an independent committee to investigate the society's past.

Butenandt had left his entire collection of papers to the society's archive, but stipulated that they remain closed to the public until 2025. Markl decided that this did not apply to members of the committee investigating the society's past. The first person granted access to the letters was science historian Robert Proctor. His findings, which appeared in 2000 as a preprint of the committee's report, were devastating. He revealed that von Verschuer had written to Butenandt disclosing that Hillmann was helping him to analyse the Auschwitz blood samples. Butenandt then asked Hillmann to destroy his (Butenandt's) documents marked "Geheime Reichssache" (top secret) before the Russians arrived. Hillmann arranged for them to be sent to Butenandt in Tübingen, but they then disappeared. We still do not know their secret.

The committee has now produced this book, edited by Wolfgang Schieder and Achim Trunk, which contains a dozen articles on various aspects of Butenandt's life. The committee is proud of the fact that it was written by historians, not scientists. Perhaps the most surprising aspect is that Proctor's article is not included. So what does the book tell us instead?

Schieder has written a piece about the best science in the Weimar Republic and the third Reich. He claims that Butenandt was a member of a nationalistic, anti-Semitic fraternity — he was no Nazi by ideology, but shared some of their views. Schieder's most startling discovery is that Butenandt was accepted as a Nazi party member on the same day in May 1936 that he was appointed as director of the Kaiser Wilhelm institute. Party membership was apparently the price he paid for the position.

Helga Satzinger writes about Butenandt's relationships with women. He accepted women in science only as technical assistants, and then only if they were single and

attractive. He went on to marry his own technical assistant, who came from a high-class family, and they had seven children together. Butenandt never supported the two female professors in his faculty at the Kaiser Wilhelm institute.

A chapter by Hans-Jörg Rheinberger deals with Butenandt's work with Kühn. Amazingly, despite their collaboration, they did not publish a single paper together. Butenandt published his research with his postdocs, but never with equal partners — he wanted all the glory for himself.

Carola Sachse writes about Butenandt's friendship with von Verschuer. And Trunk provides a new suggestion about experiments that von Verschuer considered doing with the blood samples from Auschwitz. Butenandt's proposition that von Verschuer was hoping to develop blood tests to identify different races, if true, makes his work even more sinister than was previously thought.

I think that a 450-page book about a great scientist should have at least one chapter dealing with his discoveries. What did he discover and when? Who with? What did he miss? There are a few answers here and there, but a single chapter would in my view have been a useful addition. Historians who believe that the content of science is a social construct may not mind, but scientists will.

I applaud the publication of this book, but am disappointed that it has been published only in German. Should the international community be denied the tale of this sad piece of history? ■

Benno Müller-Hill is at the Institute of Genetics, University of Cologne, Weyertal 121, Cologne 50931, Germany.

Science in culture

Dalí's immortality of the soul

A documentary film reveals the great surrealist's passion for science.

Alison Abbott

"Thinkers and literati can't give me anything. Scientists give me everything, even the immortality of the soul." Perhaps surprisingly, these are the words of surrealist painter Salvador Dalí, who died in 1989, with the books of physicists Stephen Hawking and Erwin Schrödinger, and mathematician Matila Ghyka, on his bedside table.

Dalí's paintings are almost too familiar. They include *The Great Masturbator*, a classically Freudian work, *The Persistence of Memory*, with its iconic image of clocks dripping from dead branches and flowing across surfaces, and *The Face of War* (shown here).



But few know of Dalí's lifelong passion for science, and the eagerness with which he sought the company of scientists. The documentary film *The Dalí Dimension*, which had its first international showing at the EuroScience Open Forum 2004 meeting in Stockholm, Sweden, last month, puts the record straight. Making extensive use of little-known film footage, much of which has never been aired before, Catalan film-maker Joan Úbeda documents the side of Dalí that engaged with the fundamental theories and discoveries of his time — such as relativity, quantum physics and genetics — and reflected them in his work.

In a powerful narrative, the film introduces

Salvador Dalí often incorporated science into his paintings, such as *The Face of War* (right), and can be seen below with a scientific magazine under his arm.



us to Dalí's unusual relationships with some top scientists,

including Nobel laureate Ilya Prigogine, who entered a debate with him about time and relativity. Another was James Watson, who in 1965 boldly asked him to illustrate his book *The Double Helix*, though Dalí didn't do it in the end.

Dalí held court in the New York hotel he made his second home and summoned those scientists he thought might interest him. But he was far from being a scientific dilettante. "We always spoke pretty much as equals," reports mathematician Thomas Banchoff, who was a frequent visitor.

Three years before his death, Dalí summoned a large group of top scientists to a conference on 'Chance in Nature' in the Dalí Museum in Figueres,

Spain. The artist watched the proceedings on a TV monitor from his sick-bed. At the end he called Prigogine and his good friend the mathematician René Thom to his bedroom. He asked them to reconcile — "in Schrödinger's name" — their scientific differences, which they had exposed in a quarrel during the meeting. "I did not understand why he invoked Schrödinger," comments Prigogine. But he probably understood that Dalí had enjoyed being a spectator to the dramatic battle of words.

In recognition of the centenary of Dalí's birth, a major retrospective of his work has recently opened at the Palazzo Grassi in Venice and will run until January 2005. The film, meanwhile, has its formal première in Barcelona on 16 September and will be shown by television companies in several European countries in the coming weeks.

♦ www.dalidimension.com

Defeating dementia

New Frontiers in Cognitive Aging

edited by Roger A. Dixon, Lars Bäckman & Lars-Göran Nilsson

Oxford University Press: 2004. 362 pp.

£49.95, \$99.50

George M. Martin

John Morley, the retiring editor of a leading journal of medical gerontology, wrote a farewell editorial entitled "The Top 10 Hot Topics In Aging" (*J. Gerontol. A Biol. Sci. Med. Sci.* 59, 24–33; 2004). Not surprisingly, perhaps, cognitive decline was the first topic on his list. It is probably the single most feared aspect of growing old. It may also be among the most complex and least understood features of a host of age-related changes that can be summarized as senescent phenotypes.

Fortunately for those of us on the verge

of entering the 'fourth age' (defined by Paul Baltes and Jacqui Smith as the age at which 50% of those who are alive at ages 50–60 years subsequently die), help may be on its way. There are, for example, efforts to define more accurately 'minimal cognitive impairment' and its transition, in many instances, to dementias of the Alzheimer type. In developed countries, the fourth age comes at about 80–85 years; this is when various behavioural and physiological compensatory mechanisms — what I have called 'sageing' — no longer work very well. But events that modulate the age of onset and rate of progression of senescent phenotypes can have their origins much earlier. A fuller understanding of cognitive ageing therefore requires a lifespan approach, as first appreciated by my colleagues in social gerontology.

New Frontiers in Cognitive Aging has its origins in a conference held in an exceptionally pleasant ski resort in British Columbia, Canada. The contributors include Swedish,



Zbigniew Pronaszko's painting *Golden Wedding Anniversary* shows the decline caused by old age.

Canadian, US and Australian scientists. According to the editors, both the conference and the book were designed to bring

together workers from different disciplines to consider three 'new frontiers' in cognitive ageing: new theoretical directions; integrations of findings from cognitive neuroscience with cognitive ageing; and the effects of biology and health on cognitive ageing.

The book will be of most interest to cognitive psychologists who specialize in memory research and want an update on some of the major findings from such methods as functional magnetic resonance imaging and positron emission tomography (PET). Those new to the field will have to go elsewhere for basic definitions of such terms as episodic and semantic memory. There is a little genetics thrown in, too — notably some brief discussions of the influence of polymorphic alleles at the apolipoprotein E locus on cognitive functioning and susceptibility to dementias of the Alzheimer type. We still do not know the mechanistic details of how specific alleles at this locus affect cognitive functioning even in middle age, or why these polymorphic alleles evolved and are maintained at such variable frequencies in many human populations.

We can certainly be grateful for the achievements of cognitive psychologists, epitomized by the pioneering work of Timothy Salthouse on processing speed, which is well discussed in this volume. The findings derived from joint efforts of cognitive psychologists and neuroimaging experts are also nicely documented here. A striking example from a Swedish PET study is illustrated in a chapter by Lars Nyberg and Lars Bäckman, showing the altered prefrontal lateralization associated with the retrieval of certain forms of memory in normal aged human subjects compared with younger people.

What surprised me, however, was the paucity of information about recent studies on the cellular and molecular aspects of cognition and their significance for our understanding of cognitive ageing. The efforts of Ancino Silva are particularly notable: he has set up the Molecular and Cellular Cognition Society (www.molcellcog.org), which meets in connection with the annual meeting of the Society for Neuroscience, a cosy little gathering of almost 30,000 neurobiologists.

The cover image of this book, a painting by Zbigniew Pronaszko shown on the previous page, reminded me that depressive illness is very common in older people and may sometimes be misdiagnosed as an early form of dementia. ■

George M. Martin is emeritus director of the Alzheimer's Disease Research Center, University of Washington, Seattle, Washington 98195, USA.

More on ageing

Forever Young: A Cultural History of Longevity

by Lucian Boia

Reaktion, 224 pp. £16.95, \$25



Is it a cave? The Flatiron Building in New York was the tallest skyscraper of its day.

Back to the walls

The Urban Cliff Revolution: New Findings on the Origins and Evolution of Human Habitats

by Douglas Larson, Uta Matthes, Peter E. Kelly, Jeremy Lundholm & John Gerrath

Fitzhenry & Whiteside: 2004. 272 pp. \$28.95

Clive Gamble

In uncertain times, where do we feel safest? The answer usually involves strong walls, whether they are part of a castle keep, a nuclear shelter or a panic room. We also feel that there is safety in numbers, thanks to combined strength and abundant look-outs, and in remote places, such as desert islands. But these are all short-term views. When asked by evolutionary ecologists, the question becomes: would we be better off up a tree in a tropical rainforest, out on the savannah or lurking in reed beds around a lake? The answer must take into account our evolutionary history and consider how security contributed to the radiation into so many different fossil species. It also needs to ask whether our current psychological response to different places, in terms of where we feel safe, can be traced to our evolutionary past.

A wide range of evolutionary disciplines, including archaeology, palaeoanthropology,

psychology, biology, molecular genetics and quaternary science, have tried to find answers to this question, which often leads to a search for the original environmental setting to help explain our present condition. The search is expressed in many ways but tends to encompass the garden of Eden, cradles of human civilization, evolutionary adaptation and E. O. Wilson's biophilia, the biological basis of attraction.

Ecologists add another factor in this engagingly written book. The 'urban cliff' hypothesis argues for the importance of cliffs and other rocky outcrops for a very productive group of species: humans and our commensals and mutualists. Out of these rocky habitats, which account for less than 1% of the Earth's vegetation-covered surface, come an astonishing array of plant and animal species that we have removed from these ecologically poor refuges. These pets, pests and food items are indicative for the authors of the importance of a rock-walled paradise in human evolution. To demonstrate this, they rehabilitate the caveman into a cliff-dweller and give him a mission: to recreate the ancestral home not as an urban jungle but as concrete city canyons.

Sometimes the authors' enthusiasm for their hypothesis carries them away. Their photographs liken the steps of the Capitol building in Washington to the slope in front of a cave, and a gallery in the Metropolitan Museum of Art in New York represents a modern-day version of an art-filled cave. They suggest that we are deeply attracted to cliffs because they once provided places of safety. Our liking for tower blocks stems from a residual biological memory.

But if there is a habitat comparable to the modern city, it is surely the tree tops. High above our city streets is a place of safety where a well adapted superhero such as Spiderman can put aside his psychological traumas and patrol the unsafe streets below. By comparison, the Batcave is nothing more than a glorified garage. Rock shelters, in my experience, are great if you have forgotten your tent. They retain heat in the back wall like a night storage radiator — but beware of falling rocks.

The urban cliff hypothesis is another way of saying that humans are involved in niche construction. But there never was a rock-walled paradise with good homeland security against prowling predators. Humans act on the environment just as it acts on us. The 'natural' landscape does not exist today and never has. Our psychological response to the modern world does not depend on where we went for an untroubled night's sleep a few hundred thousand years ago. ■

Clive Gamble is professor of geography at Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK, and co-director of the British Academy Centenary Project, Lucy to Language: the Archaeology of the Social Brain.

Enter transfer RNA

Theory and experiment meet to find the key adaptor for gene translation.

Mahlon Hoagland

By the 1950s, scientists generally assumed that converting genetic information into the substance of life was a matter of translation. A DNA sequence, made from a combination of four kinds of nucleotides, was translated into a protein sequence, which was made from a combination of twenty kinds of amino acids. But how did this translation occur, and what machinery was involved? These were the big questions at that time.

How the answers were found includes an unexpected subplot that involves the joining of two groups of scientists with very different backgrounds and interests — biochemists and molecular biologists. Biochemists were traditional laboratory types, concerned with dissecting the cellular machinery. Molecular biologists were a new breed who were immersed in the detailed structure of large molecules and in information processing.

By 1952, Paul Zamecnik and his fellow biochemists at Harvard's Department of Medicine had successfully broken open animal cells and separated the key components by ultracentrifugation. When in combination, two of these components were capable of carrying out protein synthesis in the test tube. These were ribosomes — cytoplasmic particles on which the final linking of amino acids to make proteins occurred — and a soluble fraction that was rich in a variety of molecules whose functions were not yet known. In addition, the nucleotide ATP was found to supply the energy essential for the process.

I joined Zamecnik's group in 1952 and began looking for evidence for the presumed initial step in protein synthesis — the energizing, or activation, of amino acids. Using techniques learnt the previous year working in the laboratories of Fritz Lipmann, a pioneer in biochemical energetics, I found that the soluble fraction was rich in a set of enzymes that attached the adenosine monophosphate part of ATP to amino acids, creating aminoacyl-AMPs. This modification provided the amino acids with the energy they would need to react with each other to form a chain. I reported these findings at a meeting organized by molecular biologists in 1955. Interest was lukewarm — the audience was more interested in how amino acids were arranged in specific sequences (how they were ordered) than in how they were energized.

In 1956, Zamecnik and his colleague Mary Stephenson made a surprising discovery: in the presence of ATP, amino acids were also attached to a small quantity of RNA found in the soluble fraction. Although most of the total cellular RNA was in the ribosomes, around 10% was found in the soluble fraction and was presumed to be 'junk' — fragments of the larger RNA from the ribosomes, perhaps produced in the process of rupturing the cells and extracting their contents. We called it soluble RNA (sRNA).

In the meantime, researchers at the

RNA templates on ribosomes. Francis circulated this 'adaptor hypothesis' among 20 fellow molecular biologists of the RNA Tie Club in 1955, but it was not formally published until 1958 (*Symp. Soc. Exp. Biol.* **12**, 138–163; 1958).

Back in Boston, unaware of Francis's brilliant imaginative leap, I was pursuing the peculiar fact that amino acids seemed to bind to an RNA component of the cellular fraction that catalysed the ATP-dependent activation of amino acids. (It was intriguing to assume that the natural function of the activating enzymes was to transfer their aminoacyl-

AMPs to these sRNA molecules. This later proved to be true.) Were these sRNA-bound amino acids essential intermediates in protein synthesis? With thumping heart, I did the key experiment. I first briefly incubated the soluble fraction with amino acids and ATP, to attach amino acids to sRNA, then removed unattached amino acids and ATP before incubating the fraction with ribosomes. To my delight, the amino acids on sRNA were rapidly transferred to their final linkage in protein on ribosomes! From that moment on, we had little doubt that sRNA (later to be renamed transfer RNA or tRNA) was the physical link between activated amino acids and their ordered arrangement in proteins (*Biochim. Biophys. Acta* **24**, 106–107; 1957).

In late 1956 Jim Watson, recently appointed to the Department of Biology at Harvard, learned of our findings through the efficient Boston–Cambridge (Massachusetts) grapevine and paid us a visit. After hearing our account of the discovery of sRNA, he asked if we knew of Francis Crick's adaptor hypothesis. Acknowledging our ignorance and somewhat miffed that a molecular biologist had foretold the existence of the intermediate we had discovered, we couldn't help but admire Francis's prescience. An image arose before me: we explorers, slashing and sweating our way through a dense jungle, rewarded at last by a vision of a beautiful temple — looking up to see Francis, on gossamer wings of theory, gleefully pointing it out to us!

And so it was that tRNAs and their companion activating enzymes (which came to be known as aminoacyl-tRNA synthetases), framed by the adaptor hypothesis, brought the classical biochemists and the molecular biologists together, snug in the same discipline, all speaking the same language. ■

Mahlon Hoagland is emeritus director of the Worcester Foundation for Biomedical Research, Worcester, Massachusetts 01605, USA.



Cavendish Laboratory in Cambridge, England — the vanguard of those who came to be known as molecular biologists — were engaged in the study of the structure of proteins and DNA. 1953 was the bright year of revelation — James Watson and Frances Crick unveiled the structure of DNA. By this time, scientists generally believed that RNA copies of single strands of DNA, acting as templates prescribing the sequences of amino acids in proteins, existed on ribosomes. Frances Crick turned his attention to how amino acids might be ordered on such presumed templates. As there is no chemical similarity or complementarity between amino acids and nucleotides, and thus no means by which they could directly interact, Crick suggested that amino acids might be first attached to short single strands of RNA nucleotides, thereby making the amino acids 'recognizable' to complementary sequences of nucleotides on the templates. In its simplest form, 20 specific enzymes would catalyse the attachment of 20 different kinds of amino acids to 20 different RNA 'adaptor' molecules. These would then be ordered by complementary nucleotide pairing on single-stranded

B. OGDON

Carbon nanotubes tune up

A. N. Cleland

Electromechanical resonators are components in many technologies. A nanometre-size version — a resonating carbon nanotube — has now been created that can be tuned over a range of frequencies.

The simplest stringed musical instrument is the bow, similar to a hunter's bow — a string stretched on a frame so that, when plucked, it resonates and produces a note whose pitch is determined by the string's tension. On page 284 of this issue, Sazonova *et al.*¹ demonstrate a nanometre-scale version of a musical bow by exciting and detecting mechanical resonances in a carbon nanotube suspended between two gold electrodes. The authors show that they can tune the resonance frequency, or pitch, of the nanotube by varying a gate voltage, which serves as both an electronic tension adjustment and a means of exciting the vibrations. The vibrational frequencies of the nanotube are more than a thousand times higher than the audible range, so the tones produced cannot be heard. However, in an elegant demonstration of the diverse capabilities of carbon nanotubes, Sazonova *et al.* have used the nanotube itself as an electronic detector — one that can 'hear' its own motion.

As well as their uses in music, mechanical resonators are key to many technologies, serving for instance as clocks and electronic filters in radio receivers and mobile phones. Quartz mechanical resonators are commonly used in thin-film processing as monitors of film thickness, capable of measuring the mass of a film that is only one atomic layer thick. Work is under way to develop similar micromechanical resonators for mass spectrometry^{2,3}, with the aim of measuring mass at the single-molecule, or even the single-atom, level. Efforts are also being made to observe and use mechanical resonators at the quantum limit^{4,5} — the extremely small-energy limit, where the energy of motion of the resonator can take on only discrete values, rather than the continuous range allowed in the higher-energy classical regime. For such applications, desirable features would be resonance frequencies in the megahertz to gigahertz (10^6 – 10^9 Hz) range, together with small mass and compact size; tunability is also a much sought-after capability.

Carbon nanotubes — with diameters of only a few nanometres, a controlled chemical

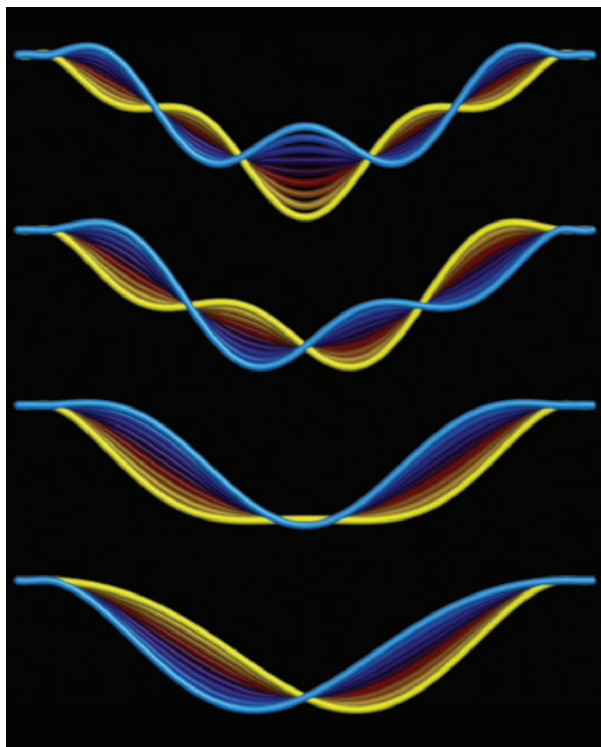


Figure 1 Good vibrations. These simulations show the first four resonant modes of a carbon nanotube, which is clamped at each end. Sazonova *et al.*¹ have detected such vibrations, and have demonstrated their tunability over a range of frequencies — like the strings of a violin.

make-up, intriguing electrical properties and extremely high strength and stiffness — may prove very useful as mechanical resonators. Their low mass and ability to withstand large tensile stresses indicate that they may have the potential to operate at very high resonance frequencies, and over an enormous tuning range. Nanotubes free of defects might also achieve very low mechanical loss, corresponding to a high mechanical quality that would be of great value in both timing and sensing applications. Much effort has therefore been devoted to exciting and detecting mechanical resonances in carbon nanotubes — but without much success. Sazonova *et al.*¹, however, at last provide a clear demonstration of the detection and tunability of both single and multiple resonances — tones and overtones — in carbon nanotubes (Fig. 1).

Detecting the motion of such structures was a stumbling-block in previous attempts because of the small sizes and high

frequencies involved. Here¹, the nanotubes' motion was detected using the variation in their electrical conductance as they moved through the electrical field of the gate electrode. However, because the nanotubes' high electrical resistance made it difficult to extract the direct signal (that at the resonance frequency), the authors used the non-linearity in this electrical response to demodulate the signal to more easily detected audio frequencies. This is analogous to what happens in a radio receiver, where a MHz-frequency carrier is demodulated to extract the superposed audio signal. Resonance frequencies ranging from 5 to 150 MHz were excited and detected. Each resonance could be tuned by varying the gate voltage, thus varying the nanotube's tension; tuning was achieved in some cases over more than one octave. And some of the nanotubes displayed at least three different resonances, corresponding to overtones of the fundamental vibration mode, all of which were externally tunable.

One disappointment in the results is the low resonance quality factor, *Q*. The highest *Q* values measured were about 200. This means

that, once a nanotube is set vibrating, half of the initial energy of motion is lost within about 200 oscillations. A tuning fork, with a significantly higher *Q* value than this, loses much of its energy by exciting sound waves in the air around it (which allows the listener to hear it). But the highest *Q* values for the nanotubes were measured in a vacuum, eliminating sound waves as a possible channel for energy loss. When the ambient pressure was increased to just under 1% of atmospheric pressure, Sazonova *et al.*¹ found that the *Q* value was significantly reduced. Whether the small *Q* — and thus the large loss — is due to contamination of the nanotube surface, to friction within its structure, or to the loss of vibrational energy in the mechanical supports, is clearly a topic for further investigation.

But Sazonova and colleagues' achievement provides a clear direction for future work: to understand the mechanical properties of carbon nanotubes; to develop

H. USTUNEL & D.

applications for these smallest of resonators; and, possibly, to further advance our understanding and control of mechanical-energy loss. Future efforts may add multi-stringed instruments to the present device — and perhaps, in time, arrive at a full symphony orchestra.

A. N. Cleland is in the Department of Physics, University of California at Santa Barbara,

Santa Barbara, California 93106, USA.

e-mail: cleland@physics.ucsb.edu

1. Sazonova, V. *et al.* *Nature* **431**, 284–287 (2004).
2. Ekinci, K. L., Huang, X. M. H. & Roukes, M. L. *Appl. Phys. Lett.* **84**, 4469–4471 (2004).
3. Ilic, B. *et al.* *J. Appl. Phys.* **95**, 3694–3703 (2004).
4. Knobel, R. G. & Cleland, A. N. *Nature* **424**, 291–293 (2003).
5. LaHaye, M. D., Buu, O., Camarota, B. & Schwab, K. C. *Science* **304**, 74–77 (2004).

Cell biology

Myosins meet microtubules

Margaret A. Titus

A central part of the machinery of cell division is the spindle. The creation and operation of this structure seem to require a component of the cell's infrastructure not previously associated with it.

A shocking thing has happened in the world of the cytoskeleton, the complex of proteins responsible for cell shape and movement. One of the most important structures it forms is the spindle, which ensures the faithful delivery of replicated chromosomes to daughter cells following cell division. Spindle assembly was once believed to be the sole responsibility of the cytoskeletal components known as microtubules, and their associated motor proteins (the dyneins and kinesins). The paper by Weber *et al.* on page 325 of this issue¹ now shows that the process depends — in some circumstances at least — on another cytoskeletal component, actin, and an associated motor protein, a member of the myosin family, which can bind directly to microtubules.

A spindle is necessary for both meiosis, the form of cell division that produces gametes for sexual reproduction, and mitosis, cell division for growth. Working with unfertilized eggs, or oocytes, of the amphibian *Xenopus*, Weber *et al.* show, surprisingly, that one of the members of the highly diverse family of myosins, myosin-10 (Myo10), is involved in both spindle assembly and the subsequent positioning of the nuclei during meiosis.

The view that the spindle depends solely on the microtubule cytoskeleton arose from experiments showing that up until its final act — the creation of daughter cells via a process called cytokinesis — cell division was a myosin-free event. Early studies^{2,3} looked at only one form of myosin, Myo2. But subsequent work⁴ demonstrated that mutants lacking other myosins (such as types 1, 5, 6, 7 and 15) can undergo mitosis perfectly well.

The first hint of direct interaction between a myosin and microtubules came from an initially puzzling observation^{5,6} — that in a certain cell type, Myo5, known as a motor that powers organelle transport, localizes to microtubules as well as to the

centrosome (a structure from which the spindle develops), in addition to being present in regions where actin resides. Consistent with these observations, it later emerged that Myo5 binds directly to microtubules *in vitro* with high affinity via its tail region⁷.

Another indication of potential microtubule–myosin interaction came from the observation that actin filaments can slide along microtubules in extracts of *Xenopus* oocytes, and that this movement is inhibited by the addition of a Myo5-binding antibody⁸. Finally, studies of the microtubule-binding domains of a plant kinesin, KCBP, revealed that a microtubule-binding domain resides in its amino terminus⁹. The region responsible was identified as the MyTH4 (myosin tail homology 4) domain, which, as its name indicates, is found in several different myosins. This finding suggested that

myosins — or, for that matter, any other protein — with a MyTH4 domain might interact directly with microtubules.

Weber *et al.*¹ now show that Myo10 from the *Xenopus* oocyte does indeed bind microtubules and that it is involved in spindle assembly and positioning of the nucleus during meiosis. The carboxy-terminal tail (or non-motor region) of Myo10 contains three domains (known as pleckstrin homology domains) that bind membrane lipids, and also a region combining both a MyTH4 and a FERM domain¹⁰. The FERM domain is always found in association with a MyTH4 domain in myosin tails and interacts with myosin-binding partners⁴. Myo10 is found only in vertebrates. It is best known for its roles in forming the extensions at the cell edge that produce motion, and in engulfing external particles or debris^{10,11}, but it may also be a player in processes involving adhesion¹².

Weber *et al.* found that *Xenopus* Myo10 co-localized with microtubules along their length in oocyte extracts and, surprisingly, that it did not seem to be significantly associated with actin. *In vitro* binding assays showed that it could also bind directly to microtubules by its carboxy-terminal MyTH4–FERM domain. Unlike the case with KCBP⁹, the MyTH4 domain alone bound only weakly: full binding required MyTH4 and FERM combined.

The meiotic nucleus of the oocyte typically abuts the actin-rich cortex at the cell periphery, where it is anchored by nuclear microtubules that extend into the cortex. Myo10 is localized at the edge of the meiotic spindle in the intact oocyte, in a position that would permit it to link the external spindle microtubules to the actin-rich cell cortex¹. Microinjection of RNA that encodes the tail

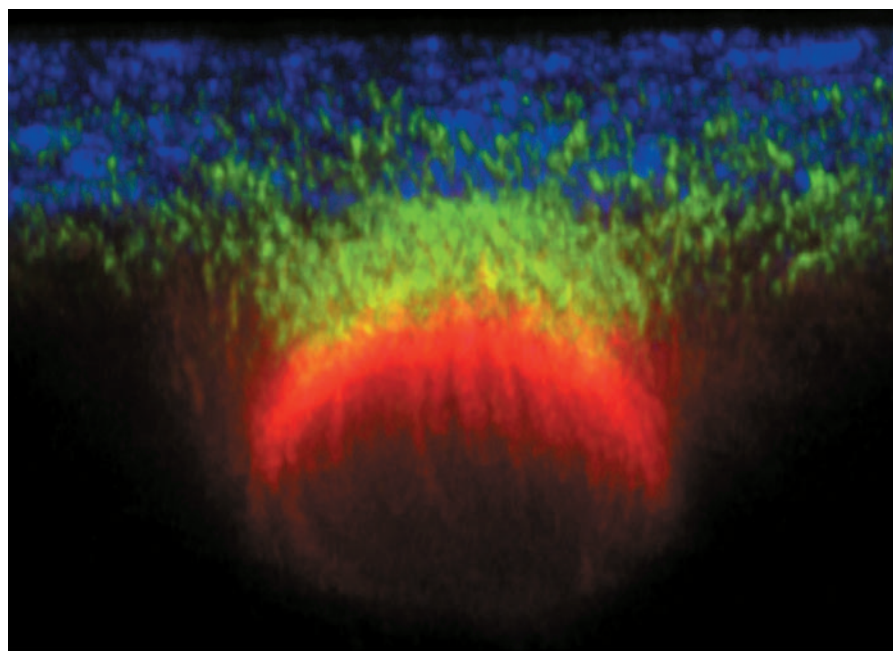


Figure 1 A myosin connection. Weber *et al.*¹ show that myosin-10 (green) links microtubules (red) and actin (blue) in the meiotic *Xenopus* oocyte.

containing the three pleckstrin homology and MyTH4-FERM domains both perturbed spindle assembly and displaced the spindle from its normal position on the cortex. Co-injection of RNA encoding the full-length Myo10 suppressed this defect; injected antibodies directed against the Myo10 motor domain resulted in a similar defect.

Weber *et al.*¹ have uncovered a direct functional link between a motor of the actin cytoskeleton and microtubules (Fig. 1). But how does that motor anchor the nucleus and contribute to spindle assembly? Like all myosins, Myo10 uses ATP to power movement along actin filaments; so Myo10 might simply provide the motive force to slide the spindle up towards the cortex by moving along polarized actin filaments that extend from the cortex into the cytoplasm. This positioning and applied force may be necessary to maintain the integrity of the spindle. Alternatively, as Myo10 may spend most of the ATP enzymatic cycle off the actin filament¹³, it could have a fine-tuning function that involves making small adjustments to keep the spindle in just the right position. Yet another possibility is that Myo10 does not act as a motor as such. Rather, it might be an ATP-sensitive crosslinker between the actin and microtubule cytoskeletons that can be modulated by other regulatory elements, with the overall interactions being required for spindle formation.

It remains to be seen whether Myo10 has a role in spindle positioning and assembly in other cell types, and whether its other functions require interaction with microtubules. And could it be that other myosins also interact with microtubules, either routinely or in specialized situations? All in all, Weber and colleagues' results¹ may in due course be seen to have turned the cytoskeletal field on its ear. For the moment, they serve as a reminder that, given the functional diversity of the myosin superfamily and the fact that many members of the family remain uncharacterized, one can never be too open-minded when considering what myosins may be involved in. ■

Margaret A. Titus is in the Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, Minnesota 55455, USA.
e-mail: titus@mail.ahc.umn.edu

1. Weber, K. L., Sokac, A. M., Berg, J. S., Cheney, R. E. & Bement, W. M. *Nature* **431**, 325–329 (2004).
2. Kiehart, D. P. *et al.* *J. Cell Biol.* **94**, 165–178 (1982).
3. Neujahr, R., Heizer, C. & Gerisch, G. *J. Cell Sci.* **110**, 123–137 (1997).
4. Kieck, M. C. & Titus, M. A. in *Molecular Motors* (ed. Schliwa, M.) 3–44 (Wiley-VCH, Weinheim, 2003).
5. Espreafico, E. M. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 8636–8641 (1998).
6. Wu, X. *et al.* *Cell Mot. Cytoskel.* **40**, 286–303 (1998).
7. Cao, T. T. *et al.* *Mol. Biol. Cell* **15**, 151–161 (2003).
8. Waterman-Storer, C. *et al.* *J. Cell Biol.* **150**, 361–376 (2000).
9. Narasimhulu, S. B. & Reddy, A. S. *Plant Cell* **10**, 957–965 (1998).
10. Berg, J. S. & Cheney, R. E. *Nature Cell Biol.* **4**, 246–250 (2002).
11. Cox, D. *et al.* *Nature Cell Biol.* **4**, 469–477 (2002).
12. Zhang, H. *et al.* *Nature Cell Biol.* **6**, 523–531 (2004).
13. Homma, K. *et al.* *J. Biol. Chem.* **276**, 34348–34354 (2001).

Geochemistry

The clock's second hand

Alex Halliday

The relative abundances of magnesium isotopes in the Allende meteorite reveal the precise chronology of the early Solar System — a geochemical second hand on the clock of creation.

To illustrate the enormity of geological time, it is sometimes helpful to compress it into a more meaningful reference frame. For example, the four-and-half billion years it took to make the Earth as we know it can be mapped onto the one week of creation recorded in the biblical book of Genesis. On this basis, *Homo sapiens* appeared only in the last ten seconds before midnight on the seventh day. It's a striking analogy. It is equally impressive that scientists have now figured out how to resolve these same truly tiny fractions of geological time — not in the recent past, but at the very beginning of our Solar System's existence. On page 275 of this issue, Bizzarro *et al.*¹ present new data on the earliest objects thought to have formed in the Solar System, from which they have deduced the timescales of events that happened four-and-half billion years ago. The resolution of their approach is less than ten parts per million — better than 50,000 years.

The source of this information — the abundance of magnesium isotopes in the Allende meteorite, which fell on Mexico in 1969 — is, despite its significance, anything but original. Allende is an unusually large and thoroughly studied example of a primitive class of meteorite called carbonaceous chondrites, which are thought to be accumulations of the dust and debris from the disk from which the Solar System was built. The debris consists of mostly small (millimetre-to centimetre-sized) objects made of silicate, oxide and metal. The careful determination of the isotopic composition of these objects inside Allende, and other chondrites, provides a remarkable window on the environment, processes and history that accompanied the earliest development of the Sun and planetary embryos.

Almost 30 years ago, Lee, Papanastassiou and Wasserburg published a seminal paper² demonstrating that the aluminium isotope ²⁶Al had once existed in Allende. ²⁶Al decays to ²⁶Mg with a half-life of 730,000 years, hence none of the original ²⁶Al exists today. What this and many other such studies have shown is an excess of the daughter isotope ²⁶Mg relative to other Mg isotopes, for which the most likely explanation is that the meteoritic material once contained ²⁶Al.

The existence of radioactive ²⁶Al and other short-lived nuclides in the early Solar System suggests a source of heat for melting planetesimals, and provides a way of deducing time differences between events. It also means that

some stellar process must have created significant amounts of ²⁶Al just before the Solar System formed. What this process was has been the subject of debate. Some have proposed that another star produced this and certain other radioactive isotopes³; others that ²⁶Al was generated by radiation from the highly energetic proto-Sun⁴. If the former proposal is correct, ²⁶Al should be well mixed in with the matter of the early Solar System. The latter model allows scope for radial gradients or differences between objects within a meteorite, reflecting not the amount of time elapsed but the admixing of components from different portions of the circumstellar disk. This latter theory is supported by models, based on astronomical observations⁴, of how such disks might behave. Bizzarro and colleagues' data, however, make more sense if the Solar System was well mixed and ²⁶Al was not produced locally.

Of the objects found inside chondrites, two have particular significance in the early history of the Solar System: calcium aluminium refractory inclusions (CAIs) and chondrules. CAIs are the earliest objects yet identified that bear clear isotopic evidence of having formed inside the Solar System (they are not presolar, because then they would be expected to be isotopically highly anomalous). In fact, the age of the Solar System is currently defined as the absolute age of CAIs, the most recent and precise estimate⁵ being 4.5672 ± 0.0006 billion years, obtained from the Efremovka meteorite. CAIs are rich in the most refractory elements of the periodic table — those that are predicted to have condensed first from the hot gas (hotter than 2,000 K) of the developing Solar System as it cooled. It seems likely, therefore, that they formed in an environment close to the newborn Sun.

From their Mg-isotope data, Bizzarro *et al.*¹ show that the CAIs from Allende formed within a very narrow time window — possibly as short as 50,000 years — and that they all had the same initial abundance of ²⁶Al. If CAI production is a localized phenomenon, linked to a site where ²⁶Al is generated, how could Allende contain objects that all formed from the same reservoir of aluminium and magnesium, at the same time, after having been scattered across hundreds of millions of kilometres of space? It will be interesting to obtain similar precise Mg-isotope data for CAIs from other chondrites. Whatever is found will place powerful new constraints on how stars and circumstellar disks function.



100 YEARS AGO

Owing to very ill-health, I have not been able to make observations of Jupiter during the last few weeks, but have been interested in receiving the results of some other observers. It appears that the great red spot is rapidly accelerating its motion, so that its longitude is decreasing, and with a continuation of this behaviour the spot will ultimately correspond with the position of the zero meridian of system ii. of Crommelin's ephemerides... The variations in the velocity of the spot during the past few years have exhibited a curious oscillation, and it will be important to watch the future developments of the object. It would be interesting to see in NATURE during ensuing months some reports from observers as to whether this singularly durable marking maintains its present rapid westerly drift.

W. F. Denning

From *Nature* 15 September 1904.

50 YEARS AGO

Obituaries: Dr. A. M. Turing, O.B.E., F.R.S.

... During his first years of research he worked on a number of subjects, including the theory of numbers and quantum mechanics, and started to build a machine for computing the Riemann ζ -function, cutting the gears for it himself. His interest in computing led him to consider what sort of processes could be carried out by a machine: he described a 'universal' machine, which, when supplied with suitable instructions, would imitate the behaviour of any other; he was thus able to give a precise definition of 'computable', and to show that there are mathematical problems the solution of which are not computable in this sense. The paper which contains these results is typical of Turing's methods: starting from first principles, and using concrete illustrations, he builds up a general and abstract argument. Many years later he used an elaboration of the same ideas to prove the unsolvability of the word problem in semi-groups with cancellation... He was awarded the O.B.E. for the work he did during the War, and after it he was invited by the National Physical Laboratory to direct the design of an electronic digital computer (which he christened "The Automatic Computing Engine")... while the final construction was in progress he turned to long-term problems, considering how machines might be made to learn by trial and error and the ways in which they could be compared with human brains.

R. O. Gandy

From *Nature* 18 September 1954.

The second class of object are chondrules, the dominant component of many chondrites⁶. The most widely accepted theory is that these once-molten spheres formed from shock waves generated in the circumstellar disk. Some chondrules have been shown to have formed about 2 million years after CAIs^{5,7}. The data presented by Bizzarro *et al.*¹ show that chondrule formation seems to span a range of time, extending to 2 million years after the birth of the Solar System. There is now excellent agreement between chronology data from uranium-lead decay⁵ and from aluminium-magnesium decay^{1,7}, and this implies that ²⁶Al was not produced locally; rather, it was uniformly distributed in the solar nebula. At least some chondrules overlap with the ages of CAIs, having formed extremely early. This has never been demonstrated so clearly before, even though there are rare examples of what are thought to be chondrules contained within CAIs^{6,8}. Chondrules with a high Al/Mg ratio seem to have formed later (consistent with the view of those petrographers who consider chondrule formation to involve a geochemical evolution towards high-Al types^{6,7}).

There are, however, some uncertainties in Bizzarro and colleagues' approach¹ that warrant a note of caution. Most of the ages presented are really model ages of bulk objects that assume that the data define a time of CAI or chondrule formation; it is possible that some CAIs or chondrules define an apparent age that is really that of an earlier protolith. Also, large corrections have been made for the isotopic fractionation (mass-dependent fluctuations that occur in the isotopic ratio through natural and laboratory processes). The basis for these corrections needs to be better understood for an unequivocal interpretation of the data.

Despite these reservations, Bizzarro and colleagues' interpretation is entirely consistent with other data, lending credence to their

assumptions. It would seem that many chondrules did indeed form very early. This does not preclude the possibility that some of them formed in collisions between planetesimals, as is the case in some models⁹. It has long been thought that planetary embryos (around 1,000 km in diameter) should have formed during the first 100,000 years of the Solar System's existence¹⁰. However, no isotopic proof of such objects at that time has yet been found. For example, the meteorites (eucrites) thought to have come from the asteroid 4 Vesta have strontium-isotopic compositions that are hard to explain unless that particular planetesimal formed more than 2 million years after the birth of the Solar System¹¹.

Chondrule chronology has long been one of the toughest challenges in geochemistry. Not only does this study by Bizzarro *et al.*¹ knock a neat hole in the problem, it also stretches the imagination. Refined experimental techniques and better observational data should allow excellent progress in understanding how the objects in chondrites were created and how dust is made and recycled in circumstellar disks. We will then be tracing the earliest evolution of solar systems as the second hand ticks.

Alex Halliday is in the Department of Earth Sciences, ETH Zentrum, Zürich CH-8092, Switzerland, and the Department of Earth Sciences, University of Oxford, Oxford OX2 3PR, UK.

e-mail: halliday@erdw.ethz.ch

1. Bizzarro, M., Baker, J. A. & Haack, H. *Nature* **431**, 275–278 (2004).
2. Lee, T., Papanastassiou, D. A. & Wasserburg, G. J. *Astrophys. J.* **322**, L107–L110 (1977).
3. Cameron, A. G. W. & Truran, J. W. *Icarus* **30**, 447–461 (1977).
4. Lee, T. *et al. Astrophys. J.* **506**, 898–912 (1998).
5. Amelin, Y. *et al. Science* **297**, 1678–1683 (2002).
6. Zanda, B. *Earth Planet. Sci. Lett.* **224**, 1–17 (2004).
7. Russell, S. S. *et al. Science* **273**, 757–762 (1996).
8. Itoh, S. & Yurimoto, H. *Nature* **423**, 728–731 (2003).
9. Sanders, I. S. in *Chondrules and the Protoplanetary Disk* (eds Hewins, R. H., Jones, R. H. & Scott, E. R. D.) 327–334 (Cambridge Univ. Press, 1996).
10. Lissauer, J. J. *Icarus* **69**, 249–265 (1987).
11. Halliday, A. N. & Porcelli, D. *Earth Planet. Sci. Lett.* **192**, 545–559 (2001).

Evolution

Affinity for arrow worms

Maximilian J. Telford

The origins of the arrow worms have long been obscure, but molecular studies are finally bringing the true evolutionary position of these beautiful marine predators into sharper focus.

The arrow worms are strikingly beautiful marine animals. Their transparent, slender bodies appear under the microscope like darting shards of glass. A close look at their head reveals a more sinister side; here, befitting their niche as predators on plankton, is the frightening set of 'spiny jaws' alluded to in their Latin name Chaetognatha (Fig. 1). Arrow worms are an ancient group of animals — their fossils

appear in sediments more than 500 million years old¹ — but, beyond their antiquity, little is known of their evolutionary origins. Two recent papers bring new molecular data to bear on this question, helping to shed light on the position of arrow worms in the animal family tree^{2,3}.

In the 160 years since Darwin described them as remarkable for "the obscurity of their affinities"⁴, arrow worms have been

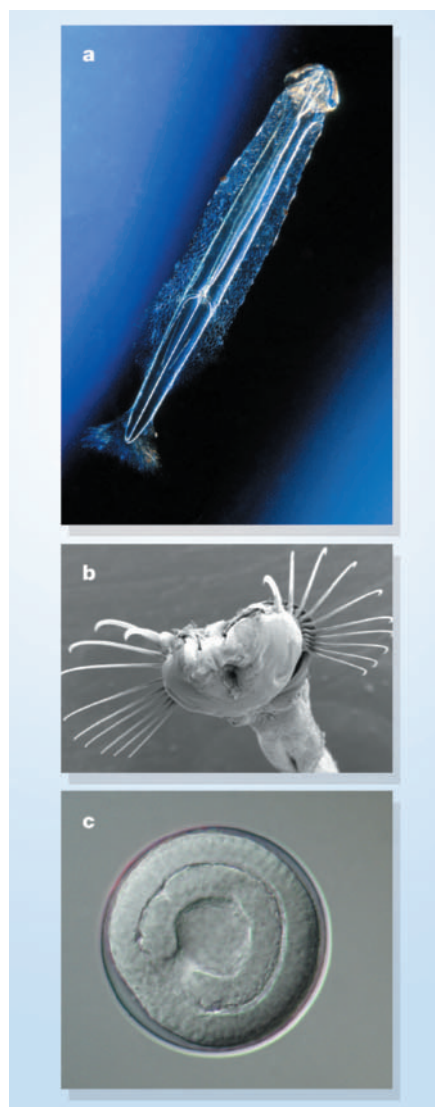


Figure 1 Beauty and the beast. a, At 4 mm long, the arrow worm (*Pterosagitta draco* is pictured here) looks like a tiny shard of glass. But its good looks hide the spiny jaws of a predator (b; *Parasagitta elegans* pictured). c, The arrow worm's mouth derives not from the first opening to form in the embryo — the blastopore, seen bottom left — but from a second opening. This characteristic is known as deuterostomy. Photographs reproduced with kind permission of: a, Peter Parks/imagequestmarine.com; b, Steven Haddock; c, David Matus.

placed by different authors in a myriad of positions in the animal kingdom. In most recent textbooks of invertebrate zoology, however, they are to be found among the last few chapters, linked to the members of a major animal subgroup called the deuterostomes, which contains the echinoderms (starfish, sea urchins and their relatives), hemichordates (acorn worms) and chordates (which include the vertebrates). During the early embryogenesis of a canonical deuterostome, the mouth, rather than being derived from the first opening in the embryo (blastopore), as it is in animals such as mol-

luscs and arthropods, forms instead from a secondary opening; indeed, deuterostome translates as 'secondary mouth'. Arrow worms too exhibit deuterostomy (Fig. 1c), and this, along with certain other embryological characteristics shared with the deuterostomes, accounts for the association.

Despite the embryological evidence, however, the one thing learned from a number of studies of arrow-worm gene sequences over the past ten years is that this deuterostome link is wrong. Wherever arrow worms do fit, it is not with the deuterostomes *sensu stricto*^{5–7}. There remain two potential general solutions to arrow-worm origins: either they emerged from somewhere within the other major animal subgroup, the protostomes ('primary mouth'), or they are a much earlier branch of the animals, their origins predating the division between protostomes and deuterostomes.

The new data come from complete sequences of the mitochondrial genomes from two closely related species of arrow worm — *Paraspadella gotoi* and *Spadella cephaloptera*^{2,3}. Mitochondria are intracellular organelles that originated as bacteria, and were taken up symbiotically by the precursor of all eukaryotes. Although almost all of the original bacterial genes have been lost from the mitochondrial genome, a stubborn few remain — 37 in most animals.

The first remarkable observation from both arrow-worm mitochondrial genomes is that they are the smallest yet known from any animal, one having 14 genes² and the other just 13 (ref. 3). Other details of the genomes differ somewhat between the two species and prevent further generalizations, but the real interest of these studies lies in their use of the mitochondrial gene sequences for determining the origins of the arrow worms within the animal kingdom. Happily, the two studies agree in positioning the arrow worms with protostomes rather than predating the protostome–deuterostome split. More frustratingly, they disagree on their exact position relative to the protostomes. This difference must arise from the different methods of analysis used by the two groups.

The *P. gotoi* sequences were analysed using a straightforward parsimony-based approach — choosing the tree that requires the fewest character changes to account for the observed gene sequences. The best tree that Helfenbein *et al.*² produce has very strong support for grouping the arrow worms as a branch next to, but outside, the protostomes.

The analyses of the *S. cephaloptera* sequences are more complex for two reasons. First, Papillon *et al.*³ use several methods to reconstruct their protein sequence tree. Second, they exploit additional sources of evidence from their mitochondrial genome. All of their analyses, apart from that using

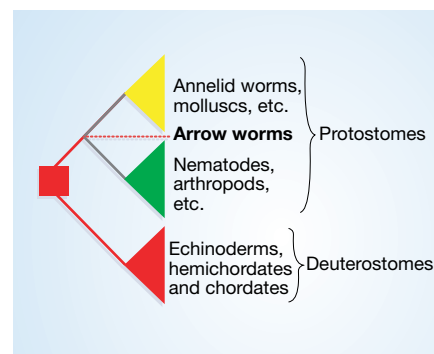


Figure 2 Evolutionary origins of the arrow worms. Recent molecular analyses^{2,3} suggest that arrow worms are related to the protostomes. Both arrow worms and deuterostomes exhibit deuterostomy and this must have been inherited from their common ancestor, represented by the red box. This ancestor also gave rise to the protostomes.

parsimony, place the arrow worms within the protostomes, associated with a subgroup that includes molluscs and annelid worms. They are cautious in their interpretation, however, and conclude only that arrow worms are linked in some way to the protostomes³.

The two additional sources of evidence used by Papillon *et al.* were comparisons of the order of the mitochondrial genes (which are probably unreliable at this level of divergence, as even these closely related arrow worms differ) and so-called 'signature sequences'. In the *S. cephaloptera* gene *NAD5* the authors find nine such signature amino acids. These are amino acids that are shared by arrow worms and by protostomes, and which differ from the equivalent deuterostome amino acids. Because the amino acids seen in the deuterostomes are also seen in 'primitive' animal species such as sea anemones, these must represent a primitive condition. The corollary is that the amino acids shared by arrow worms and protostomes must be novel characters, providing convincing evidence of a close relationship. These amino acids are in effect diagnostic of protostomes in the same way that feathers are diagnostic of birds.

To understand the significance of grouping arrow worms with protostomes, we return to the characteristics that arrow worms share with deuterostomes. Ignoring the possibility of convergent evolution, arrow worms and deuterostomes must have inherited their deuterostomy and other embryological similarities from a common ancestor (Fig. 2). The new positioning of the arrow worms means that this common ancestor also gave rise to the protostomes. This gives us an important insight into the early evolution of the protostomes, because it implies that they must have evolved from a creature that had deuterostome-like aspects of embryogenesis. This result also leaves us with a problem of nomenclature, given that

some of the protostomes are, embryologically speaking, deuterostomes.

So why are the arrow worms turning out to be so hard to place within the animal kingdom? The short answer seems to be that the arrow-worm genes studied so far have been subject to more rapid evolutionary change than those of other species⁷. The result is that, on an evolutionary tree, arrow worms form a long branch, and such long-branch species are notoriously problematic. Unfortunately, the simplest solution to the long-branch problem—finding a species of arrow worm with a more normal rate of evolution—seems impossible, because all of the 100 or so living species appear to suffer from the same problem⁸. The unavoidable conclusion

is that yet more genetic data need to be gathered in the expectation that the picture will continue to become clearer.

Maximilian J. Telford is in the Department of Biology, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK. e-mail: m.telford@ucl.ac.uk

1. Chen, J. Y. & Huang, D. Y. *Science* **298**, 187 (2002).
2. Helfenbein, K. G., Fourcade, H. M., Vanjani, R. G. & Boore, J. L. *Proc. Natl Acad. Sci. USA* **101**, 10639–10643 (2004).
3. Papillon, D., Perez, Y., Caubit, X. & Le Parco, Y. *Mol. Biol. Evol.* doi:10.1093/molbev/msh229 (2004).
4. Darwin, C. *Ann. Mag. Nat. Hist.* **13**, 1–6 (1844).
5. Telford, M. J. & Holland, P. W. H. *Mol. Biol. Evol.* **10**, 660–676 (1993).
6. Wada, H. & Satoh, N. *Proc. Natl Acad. Sci. USA* **91**, 1801–1804 (1994).
7. Halanych, K. M. *Syst. Biol.* **45**, 223–246 (1996).
8. Telford, M. J. & Holland, P. W. H. *J. Mol. Evol.* **44**, 135–144 (1996).

Physical chemistry

Quantum mechanics for plants

Graham R. Fleming and Gregory D. Scholes

To what extent do photosynthetic organisms use quantum mechanics to optimize the capture and distribution of light? Answers are emerging from the examination of energy transfer at the submolecular scale.

The first law of photosynthetic economics is: “A photon saved is a photon earned.” Research into the factors behind this principle has been burgeoning, and has recently culminated in a paper in *Physical Review Letters* by Jang *et al.*¹ in which the authors look at photosynthetic energy transfer at the quantum level.

Plants use solar antennae to capture incident photons and transmit the excitation energy to reaction centres, where it is used to initiate the primary electron transfer reactions of photosynthesis. These antennae are one of nature’s supreme examples of nanoscale engineering, and are constructed from specialized light-harvesting complexes formed of proteins that bind chlorophylls and carotenoids. Photon collection involves up to several hundred light-absorbing molecules, or chromophores. Hundreds of energy-transfer steps over a hierarchy of time scales and distances, which often occur with near-perfect efficiency², are therefore required to collect and trap solar energy.

More than 50 years ago, Theodore Förster described a method for calculating the rate of energy transfer between molecules from the overlap of the donor molecule’s fluorescence spectrum and the acceptor molecule’s absorption spectrum^{3,4}. The theory has had an enormous impact on biology, chemistry and physics. Collectively, high-resolution structural models, ultrafast spectroscopy and quantum chemical calculations have helped to expose the complex and, in some cases, subtle relationships between structure and light-harvesting in photosynthetic systems. Indeed, it has turned out that there are only a

few cases in which the energy transfer within photosynthetic light-harvesting complexes can be correctly characterized by conventional Förster theory. Moreover, the realization that the concepts elucidated during the study of light-harvesting proteins are general

principles that operate in molecular aggregates has inspired a change in thinking about energy transfer in confined geometries⁵.

Various studies^{2,6–8} have revealed that the nanoscale dimensions of the photosynthetic complexes are critical for light harvesting. Unlike most model systems, where the chromophores are spaced at distances that are large with respect to their size, chromophores in light-harvesting systems are densely packed. This means that the electronic interactions between the light-absorbing components are both qualitatively and quantitatively different from most other model systems. This realization has spurred the development of new theoretical methods, including Jang and colleagues’ work¹, which provides a detailed prescription for calculating energy transfer in multi-chromophoric assemblies.

To understand the dynamics of light-harvesting and light-trapping in photosynthesis, certain design features must be taken into account. For example, the distances between the molecules are often smaller than the overall size of each molecule. In this confined geometry, energy transfer is governed by how the donor ‘sees’ the acceptor on the submolecular scale at which the fine differences in the shape of the wavefunctions between the ground and excited states at the donor–acceptor junction become significant (Fig. 1). At this level of spatial confinement, transitions and energy levels that would be ineffective, or even inoperative, in

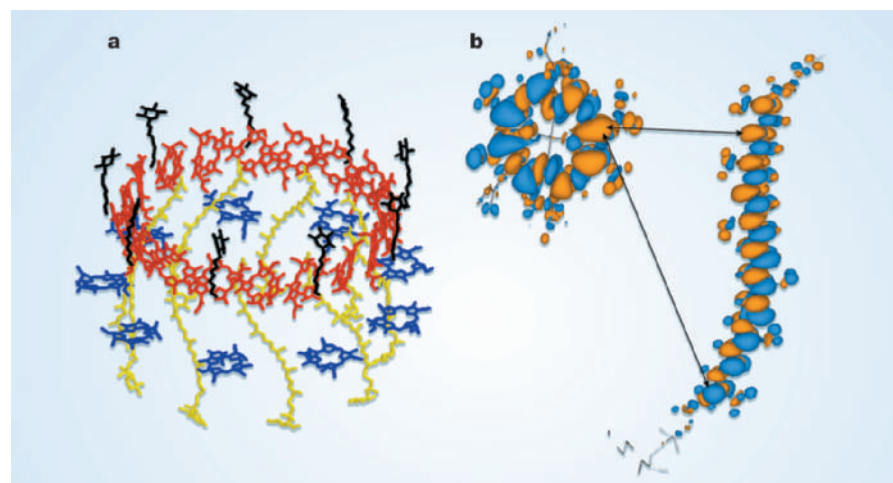


Figure 1 Designs for energy transfer. a, Chromophores in a model of light-harvesting complex (LH) 2 from the bacterium *Rhodospseudomonas acidophila* (Fig. 2), radius 3.4 nm. B800 bacteriochlorophyll molecules (blue) are widely spaced and constitute simple donors, but the B850 molecules (red) interact strongly and constitute a complex acceptor in a confined geometry. Through such interactions between molecules, photosynthetic organisms employ quantum mechanics to funnel absorbed photons to the reaction centre. On time scales of less than 1 picosecond, energy flows from the 800-nm-absorbing B800 molecules to the 850-nm-absorbing B850 molecules, and from the carotenoids (yellow) to both B800 and B850. b, A real-space picture of electronic interactions between molecules on a submolecular scale, as seen in the transition densities of LH2 bacteriochlorophyll (left) and carotenoid (right) molecules calculated from ground- and excited-state wavefunctions. The different colours represent the sign of the electron density. Instead of one average separation between donor and acceptor defining the energy transfer rate, as in Förster theory, there are clearly many length scales (examples arrowed) over which the various parts of the donor and acceptor electron densities interact.

systems with widely spaced chromophores may be crucial in increasing energy-transfer efficiency.

Typically, a confined multichromophoric system allows energy to be moved more rapidly and effectively using a broad window of acceptor states. As a result, new definitions are needed to characterize the donor and the acceptor. Physically, they each consist of multiple chromophores; spectroscopically, they behave as electronically delocalized states. The properties of these states are strongly modified by interactions with the chromophores' environments, making each donor-acceptor pair in the ensemble unique. That concept is familiar from single-molecule experiments⁹, where the spectroscopy of an individual molecule is influenced by its local environment. A general consequence is that absorption and fluorescence measurements cannot be used to record a meaningful measure of the energy transfer rate, as in the spectral overlaps that constitute the backbone of Förster theory.

These realizations led to the development of generalized Förster theory, which has helped to resolve several long-standing mysteries in photosynthesis. Paradoxically, for instance, it has emerged that the same interactions that produce perfectly efficient energy transport also allow photosynthetic organisms to construct molecular safety valves that dissipate excess excitation energy that would otherwise cause irreversible damage^{10,11}. Furthermore, this work has shown how photosynthetic systems exploit energetic disorder to improve spectral coverage, and reduce energy mismatches to make the system exceedingly robust against thermal and structural variations.

Jang *et al.*¹ have further explored the multi-component donor-acceptor aspect of natural antennae, and report an improved method of calculating the spectral overlap factors in the generalized Förster theory for predicting the rate of energy transfer in complex multichromophoric systems. They consider an additional ingredient: a kind of interchromophore term that accounts for quantum mechanical coherence among the chromophores that collectively donate, or accept, excitation energy. Such an effect, if not destroyed by environmental fluctuations on times shorter than that on which the energy transfer occurs, could lead to increased rates of energy flow. However, it is still assumed that such a quantum effect is insignificant in the donor-acceptor interaction.

In the context of this theory, we can now determine whether coherent nuclear fluctuations throughout a protein or synthetic structure can be used to accelerate the rate of energy transfer. A signature of coherence effects is that the entity acting as a donor or an acceptor will effectively change size as the amplitude of the environmental fluctuations changes with a change in temperature. Such



Figure 2 *Rhodospseudomonas* — a bacterium that makes light work.

coherence effects are not deployed in the normal operation of the best-characterized photosynthetic antenna, the light-harvesting complex 2 of the bacterium *Rhodospseudomonas acidophila*. But they could play a part in a system with more extensive delocalization among the chromophores or at very low temperature.

The work of Jang *et al.*¹ of course leaves plenty of scope for further investigations.

Molecular biology

Genetic code seizes pyrrolysine

Paul Schimmel and Kirk Beebe

Identification of the enzyme that mediates insertion of a rare amino acid, pyrrolysine, into protein solves a puzzle and expands the rules of the genetic code established nearly half a century ago.

The discovery of the genetic code that translates the information contained in DNA into proteins ranks as one of the greatest achievements of the past century. The code appeared before the split of the tree of life into the three great kingdoms — Bacteria, Archaea and Eukarya — and has been maintained in all organisms as an algorithm that relates nucleotide triplets in genes to specific amino acids in proteins. The code requires for its implementation a set of 20 universal enzymes — aminoacyl tRNA synthetases (one for each amino acid). But on page 333 of this issue, Blight *et al.*¹ show that, in certain archaea, an unusual, natural amino acid is read by the code using a novel twenty-first aminoacyl tRNA synthetase.

The adaptors responsible for translating the genetic code from the messenger RNA (mRNA) into amino-acid sequence are the

transfer RNAs (tRNAs). Each tRNA bears an 'anti-codon' sequence that corresponds to the nucleotide triplet code for a specific amino acid. Each of the canonical 20 amino acids is covalently linked to the tRNA that bears the anti-codon corresponding to the particular amino acid. This charging reaction, an aminoacylation reaction, is carried out by the aminoacyl tRNA synthetases, each synthetase being specific for a particular amino acid and tRNA. Essentially, the tRNA reads the position in the mRNA sequence where its cognate amino acid is supposed to appear, and holds the amino acid in position until it is incorporated into the growing peptide chain. Three nucleotide triplets — UAA, UAG and UGA — are used as stop signals.

For example, what is the interplay between the length and time scales over which coherent nuclear fluctuations of chromophores and their surrounding protein are preserved? Can experiments be designed to reveal these quantum effects? And on the practical side, how can these factors lead to design strategies for synthetic light harvesters with enhanced energy-transfer rates and efficiencies?

Graham R. Fleming is in the Department of Chemistry, University of California, Berkeley, and at the Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

e-mail: grfleming@lbl.gov

Gregory D. Scholes is in the Department of Chemistry, University of Toronto, Lash Miller Chemical Laboratories, 80 St George Street, Toronto, Ontario M5S 3H6, Canada.

e-mail: gscholes@chem.utoronto.ca

1. Jang, S., Newton, M. D. & Silbey, R. J. *Phys. Rev. Lett.* **92**, 218301 (2004).
2. Sundström, V., Pullerits, T. & van Grondelle, R. *J. Phys. Chem. B* **103**, 2327–2346 (1999).
3. Förster, Th. *Ann. Phys.* **2**, 55–75 (1948).
4. Van Der Meer, B. W., Coker, G. & Chen, S.-Y. S. *Resonance Energy Transfer: Theory and Data* (VCH, New York, 1994).
5. Scholes, G. D., Jordanides, X. J. & Fleming, G. R. *J. Phys. Chem. B* **105**, 1640–1651 (2001).
6. van Amerongen, H., Valkunas, L. & van Grondelle, R. *Photosynthetic Excitons* (World Scientific, Singapore, 2000).
7. Sumi, H. J. *J. Phys. Chem. B* **103**, 252–260 (1999).
8. Scholes, G. D. & Fleming, G. R. *J. Phys. Chem. B* **104**, 1853–1868 (2000).
9. Ambrose, W. P. & Moerner, W. E. *Nature* **349**, 225–227 (1991).
10. Holt, N. E., Fleming, G. R. & Niyogi, K. K. *Biochemistry* **43**, 8281–8289 (2004).
11. Jordanides, X. J., Scholes, G. D., Shapley, W. A., Reimers, J. R. & Fleming, G. R. *J. Phys. Chem. B* **108**, 1753–1765 (2004).

aminoacyl tRNA synthetase–tRNA pairs that could, in turn, insert novel amino acids into polypeptides, to create proteins with novel properties². The usual approach requires engineering a tRNA to have an anti-codon that corresponds to a stop codon, so that instead of signalling a stop, the tRNA will insert an artificial amino acid. But although these studies demonstrate the power of artificial synthetase–tRNA pairs for introducing novel amino acids into proteins, no natural examples were known to exist.

The novelty of the work by Blight *et al.*¹ rests on the identification of a natural twenty-first synthetase that directly activates pyrrolysine (both *in vivo* and *in vitro*). Pyrrolysine (pLys) is a non-canonical amino acid that is found in certain methyltransferase enzymes of archaea^{3,4}, where it is incorporated at a UAG codon, instead of there being a stop. The typical tRNA-charging reaction occurs in two steps: activation of the amino acid with adenosine triphosphate (ATP) to form an aminoacyl adenylate, and reaction of the adenylate with the tRNA to link the two. Blight *et al.*¹ show that pyrrolysine tRNA synthetase activated pyrrolysine (to form pyrrolysyl-AMP) *in vitro*, and did not activate any of the 20 canonical amino acids, including lysine. They also demonstrate that, both *in vitro* and *in vivo*, the enzyme catalyses the overall aminoacylation reaction, requiring a pyrrolysine-specific tRNA (tRNA^{pLys}), which has an anti-codon that is complementary to UAG. Recently, Polycarpo *et al.*⁵ reported similar findings *in vitro* (but not *in vivo*). Thus, a natural synthetase–tRNA pair brings a new amino acid into the genetic code.

Next, the authors¹ expressed the pyrrolysine-specific synthetase and tRNA^{pLys} in the bacterium *Escherichia coli* together with methyltransferase from the archaean *Methanosarcina barkeri*, a protein that naturally contains pyrrolysine. (Neither the two proteins nor tRNA^{pLys} are naturally present in *E. coli*.) Remarkably, when pyrrolysine was added to the culture, it was incorporated into the methyltransferase at the expected position. Thus, all the information required to expand the genetic code is contained in tRNA^{pLys}, the pyrrolysine-specific synthetase and the mRNA encoding the *M. barkeri* methyltransferase. No other components from the archaea are needed for the reaction to work in *E. coli*.

Another surprise from Blight and colleagues' work is that insertion of pyrrolysine into the growing methyltransferase polypeptide *in vivo* is highly efficient (greater than 75% efficiency), much more so than is usually seen with the artificial tRNA–synthetase pairs that act at the UAG codon (typically less than 20% efficiency)⁶. (The authors do not report any experiment to test whether tRNA^{pLys} can also insert pyrrolysine at an artificially introduced UAG codon of a non-

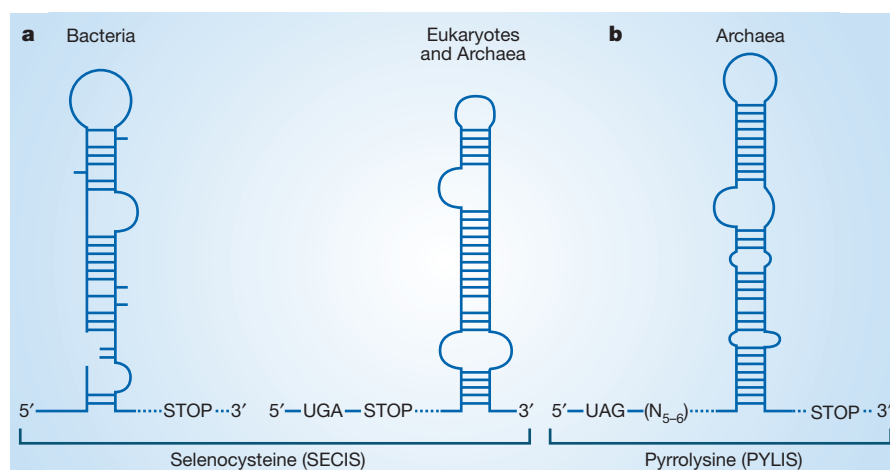


Figure 1 RNA structures signal incorporation of non-canonical amino acids. **a**, A hairpin structure called the 'selenocysteine insertion element' (SECIS) is present in the messenger RNA that corresponds to where selenocysteine is due to be inserted into the growing peptide chain. In bacteria, SelB binds to tRNA loaded with selenocysteine and to SECIS, so that instead of UGA signalling a stop, selenocysteine is added into the peptide. (In eukaryotes and archaea, two proteins perform the SelB function.) **b**, Predicted pyrrolysine insertion elements (PYLIS) in the mRNA encoding archaeal methyltransferase. These are proposed to signal for pyrrolysine to be inserted in the methyltransferase peptide at the UAG codon (also usually a stop codon) in a manner similar to SECIS. Structures adapted from Namy *et al.*⁹.

methyltransferase mRNA.) However, were tRNA^{pLys} to act efficiently at the many naturally occurring UAG stop codons in *E. coli*, the result would be expected to be toxic (which was not seen). This implies that tRNA^{pLys} can somehow 'find' specific UAG sequences.

We imagine that the local mRNA structure around the target UAG is arranged in a way that tRNA^{pLys} can recognize. A possible precedent is the mechanism by which selenocysteine is inserted into proteins. This rare amino acid is formed by a modification of the canonical amino acid serine while it is attached to a special tRNA⁷. (By analogy, an early proposal for how pyrrolysine was inserted involved a similar modification of lysine^{4,8}.) In bacteria, the selenocysteine-encoding UGA codon is targeted by a special elongation factor, SelB, that binds specifically to the selenocysteine-loaded tRNA. SelB also binds to a special RNA hairpin loop (the 'selenocysteine insertion element') that signals where the selenocysteine should go⁷ (Fig. 1a).

How does the pyrrolysine-loaded tRNA^{pLys} identify the pyrrolysine-specific UAG codons in methyltransferase RNAs? Interestingly, RNA hairpin loops similar to the selenocysteine insertion elements are predicted to lie near the pyrrolysine-encoding UAG codons⁹ (Fig. 1b). But with no other factors from methanogenic archaea expressed in *E. coli*, the possibility of a SelB-like component in *E. coli* has to be considered. Could that factor be SelB itself? The problem is that eukaryotes and probably archaea require two factors for selenocysteine insertion: a SelB-like elongation factor, and a second factor for binding the selenocysteine insertion element that has no

counterpart in *E. coli*^{7,10}. Could pyrrolysine-specific synthetase have a role in the selection of not only tRNA^{pLys}, but also the pyrrolysine insertion site, perhaps not releasing pyrrolysine-loaded tRNA until it has reached the correct position?

Although many questions remain, the main finding is that the non-canonical amino acid pyrrolysine can be translated by the genetic code using a natural synthetase–tRNA pair. In some ways, this discovery reaffirms that notion that, if something can be done by experiment, it has probably been done earlier by nature. As an example, the demonstration that RNAs could be experimentally selected to bind specific small molecules foretold the eventual discovery of such RNA 'aptamers' in bacteria¹¹. With genomes and proteomes of so many species yet to be analysed, one wonders what other surprises the genetic code has in store.

Paul Schimmel and Kirk Beebe are at The Skaggs Institute for Chemical Biology, The Scripps Research Institute, Beckman Center, BCC379, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. e-mail: schimmel@scripps.edu

1. Blight, S. K. *et al.* *Nature* **431**, 333–335 (2004).
2. Thorson, J. S. *et al.* *Methods Mol. Biol.* **77**, 43–73 (1998).
3. Hao, B. *et al.* *Science* **296**, 1462–1466 (2002).
4. Srinivasan, G., James, C. M. & Krzycki, J. A. *Science* **296**, 1459–1462 (2002).
5. Polycarpo, C. *et al.* *Proc. Natl Acad. Sci. USA* **101**, 12450–12454 (2004).
6. Anderson, J. C. & Schultz, P. G. *Biochemistry* **42**, 9598–9608 (2003).
7. Hatfield, D. L. & Gladyshev, V. N. *Mol. Cell. Biol.* **22**, 3565–3576 (2002).
8. Polycarpo, C. *et al.* *Mol. Cell* **12**, 287–294 (2003).
9. Namy, O., Roussel, J. P., Napthine, S. & Brierley, I. *Mol. Cell* **13**, 157–168 (2004).
10. Rother, M., Wilting, R., Commans, S. & Böck, A. *J. Mol. Biol.* **299**, 351–358 (2000).
11. Mandal, M. & Breaker, R. R. *Nature Rev. Mol. Cell Biol.* **5**, 451–463 (2004).

Materials science

Nanotubes get hard under pressure

Proc. Natl Acad. Sci. USA
doi:10.1073/pnas.0405877101 (2004)

When Zhongwu Wang *et al.* squeezed carbon nanotubes in a diamond anvil cell, they made something so hard that it cracked the diamond teeth of the apparatus. But they hadn't simply turned nanotubes into diamond itself: the carbon material formed under compression at room temperature seems to be different from any previously known.

The material is about 90% sp^3 carbon (the tetrahedrally coordinated form found in diamond), with a hexagonal unit cell; but Raman spectroscopy shows that it is not the well-known hexagonal form of diamond, called lonsdaleite. Rather, it seems to be a new high-pressure phase of carbon, with a density, hardness and bulk modulus (the inverse of compressibility) at least as high as those of diamond. And crucially, the new phase, which appears at pressures of around 75 gigapascals, remains stable when the pressure is released, which bodes well for technological applications of this superhard material.

Philip Ball

Information theory

Say it with black holes

Phys. Rev. Lett. **93**, 100501 (2004)

If you've really got a lot to say, the fastest way to say it is to send a package that is on the verge of collapsing into a black hole. That's what Seth Lloyd and colleagues conclude by calculating the maximum rate of communication allowed by the known laws of physics.

They start from an expression derived by Jacob Bekenstein for the maximum amount of information that can be stored in a spherical volume of space. Special relativity stipulates that this information can't be sent faster than the speed of light. But these limits don't by themselves restrict the rate at which information can be relayed, even using a finite power, because in principle the rate can be increased indefinitely by writing it in ever-denser slabs of material. The trouble is, if the slabs get too dense, they collapse into black holes.

Whether energy can be retrieved from a black hole is not known — that depends on the nature of quantum gravity. But Lloyd *et al.* assume that gravitational collapse of the information carriers is avoided — in which case the communication rate is maximized for parcels just below the Schwarzschild density for black-hole formation. But what happens when you attach the stamp?

Philip Ball

Ring cycle

The Cat's Eye nebula, captured in this image from the Advanced Camera for Surveys on the Hubble Space Telescope, is shrouded in rings of dust. But it's not clear how the 11 or so rings formed. They might have been created by a series of pulses, at 1,500-year intervals, caused by the cycle of magnetic activity on the dying star at the centre of the nebula. Alternatively, the star might have ejected a steady stream of material that has since been swept into rings. As this ring phenomenon is a common feature of nebulae, the answer may soon be at hand.

Alison Wright



Genetics

Sweet success

Hum. Mol. Genet. **13**, 1979–1988 (2004)

α -Mannosidosis is a genetic disorder that causes progressive mental and physical deterioration. Sufferers possess a defective version of the enzyme α -mannosidase, which normally helps to break down complex sugars in a cellular compartment called the lysosome. In patients, sugars build up in the lysosomes, causing them to swell and impair cell function.

Diego Prieto Rocas *et al.* have now devised an enzyme-replacement therapy that works well in a mouse model of this disorder. The authors gave the animals a single intravenous injection of human, cow or mouse α -mannosidase. A few days later, the levels of complex sugars in lysosomes had dropped significantly.

The effects were transient, however, so the authors then treated mice with two shots of the human enzyme, several days apart. One week later, the liver, kidneys and heart were completely clear of unwanted lysosomal sugars, and the levels of these sugars in the brain had dropped by around 75%. Longer-term mouse studies, with repeat injections and behavioural tests, will be the next step.

Helen Pilcher

Structural biology

Moving ammonia

Science **305**, 1587–1594 (2004)

Nitrogen is one of life's essentials, and for many organisms the main source is ammonia or closely related chemical forms. The cell-membrane controllers of ammonia transport are members of the Amt group of proteins, the structure of one of which — AmtB, from the bacterium *Escherichia coli* — has now been solved by Shahram Khademi *et al.* to 1.35 Å resolution.

One of the authors' conclusions is conceptual: that AmtB should be considered as a channel, and not a transporter. Another is that their data are consistent with a conductance mechanism that starts with

the entry of ammonium (NH_4^+) or ammonia (NH_3) into a 'vestibule', followed by pH-regulated deprotonation of the NH_4^+ , and passage of NH_3 into the cell through a 20-Å-long hydrophobic channel.

The Amt proteins are members of an ancient and widely distributed family. In humans, other members of the family include the rhesus proteins, which occur in the membranes of red blood cells and are commonly known for their significance in blood-typing. Only recently, it emerged that rhesus proteins are involved in ammonia transport: as Khademi *et al.* point out, mapping their sequences onto the AmtB structure should help in understanding rhesus function.

Tim Lincoln

Ecology

Rainforest in drought

Oecologia **141**, 114–120 (2004)

The El Niño Southern Oscillation (ENSO) — a periodic warming of ocean waters in the Pacific — has widespread climatic effects. As a result of the particularly severe ENSO event of 1997–98, southeast Asia experienced an intense drought. This affected rainforest composition, as research by J. W. F. Slik now shows.

Slik counted the number of dead trees during the drought in East Kalimantan (Borneo), Indonesia. Mortality was 11.2% higher than normal in undisturbed forest and 22.7% higher in recently logged forest. Much of this was due to high death rates (about 65%) among pioneer *Macaranga* trees, which are usually present at higher densities in managed forests, where canopy trees are more widely spaced.

This picture is complicated by the fact that increased light levels resulting from the death of canopy trees favours young, fast-growing *Macaranga* seedlings, which increased in density between 30- and 300-fold shortly after the drought. Drought may therefore indirectly favour fast-growing species — but not for long, Slik speculates. When he looked four years later, the canopy had closed up again.

Michael Hopkin

Adaptive developmental plasticity in snakes

Genes and environment stretch snake jaws to meet the demands of prey size.

The morphology of organisms is generally well matched to their environment, presumably because expression of their genes is tailored either at the population or the individual level to suit local conditions: for example, snake populations that persistently encounter large prey may accumulate gene mutations that specify a large head size, or head growth may be increased in individual snakes to meet local demands (adaptive developmental plasticity)¹. Here we test the relative contributions of genetics and environment to the jaw sizes of two tiger snake populations: one that consumes small prey on the mainland, and an island population that relies on larger prey and has a larger jaw size. Although the idea of adaptive plasticity in response to environmental pressures is controversial², we find that both factors influence the difference in jaw size between the two populations, and the influence of developmental plasticity is greater in the island population.

Snakes are ideally suited to our investigation because their jaw length constrains their maximal ingestible prey size and so reflects adaptation to local prey resources³. Tiger snakes (*Notechis scutatus*, Elapidae; Fig. 1) are viviparous Australian snakes⁴. We studied one mainland population (at Herdsman's Lake: 31° 55' S, 115° 48' E) and one island population 25 km away (Carnac Island: 32° 07' S, 115° 39' E). Adult mainland snakes feed on frogs and mice, whereas adult



Figure 1 On an island in southwestern Australia, tiger snakes feed mostly on the large chicks of silver gulls, rather than on the small frogs preyed upon by mainland populations.

island snakes mostly take silver-gull chicks (mean prey masses, 12 g compared with 30 g; circumference, 33 mm compared with 78 mm)⁶. In keeping with this dietary divergence, the island snakes grow larger and have larger heads relative to their body length^{5,6}.

To investigate whether this geographic divergence in relative head size is driven by genes that encode for head size or by the exposure of growing snakes to larger prey, we captured pregnant females in early summer and maintained them under laboratory conditions until they gave birth in the autumn. They were housed individually with access to water and a heat source, and given laboratory mice as food. Nine island females and thirteen mainland females gave birth to 123 and 129 neonates, respectively. We randomly selected one or two neonates from each litter and raised them under the same conditions as their mothers.

Half the neonates received large mice as prey, whereas the others were given the same total mass of smaller mice as prey: that is, mice fed to the two treatment groups differed in average prey mass ($F_{3,81} = 101.8$, $P < 0.001$) and circumference ($F_{3,81} = 101.7$, $P < 0.001$). The amount of food provided monthly was similar among the four groups of young snakes (island and mainland snakes, each fed either small or large prey; $F_{3,75} = 0.29$, $P = 0.84$), and led to similar growth trajectories in body mass ($F_{4,104} = 1.36$, $P = 0.25$) and snout-vent length ($F_{3,78} = 0.09$, $P = 0.96$). All snakes were measured for jaw and skull length on four occasions over 8 months using digital callipers (± 0.1 mm). The standardized husbandry conditions and split-clutch design should have minimized any maternal effects on offspring traits.

Head sizes were significantly plastic in Carnac Island juveniles with respect to skull

length (results not shown; repeated measures MANOVA; Wilks' $\lambda = 0.41$, $P = 0.03$; interaction of time and prey size, $F_{3,45} = 3.10$, $P < 0.036$) and jaw length (Wilks' $\lambda = 0.25$, $P < 0.0015$; interaction $F_{3,45} = 19.15$, $P < 0.0001$); these effects were not evident in mainland juveniles (all $P > 0.24$) (Fig. 2). Exposure to larger prey therefore seems to increase the head size of the island snakes more than that of the mainland snakes (two-way ANOVA, with origin and treatment as factors and the successive measurements of jaw length as the repeated measure: interaction $F_{3,78} = 5.66$, $P < 0.0015$) (Fig. 2).

These results indicate that adaptation could be influenced both by genes encoding for head size and by developmental plasticity. First, neonates of island snakes had larger heads than mainland snakes, despite similar body sizes (ANCOVA with location as the factor, log body-length as the covariate, log jaw-length as the dependent variable: $F_{1,244} = 25.65$, $P < 0.001$). This geographic divergence is apparent at birth, so presumably reflects hard-wired genetic differences.

Second, the relative jaw sizes of island snakes fed on large prey rapidly increased relative to the jaws of their siblings fed on small prey, whereas the mainland snakes showed no such effect (Fig. 2). We conclude that island tiger snakes can consume larger prey than mainland conspecifics for two reasons: first, they may carry genes that determine a larger relative head size; and second, their head sizes enlarge facultatively if they eat large prey.

Our study demonstrates an important ecological role for adaptive plasticity, and highlights the impossibility of dividing phenotypic variation into simplistic categories of 'nature' and 'nurture'⁷. Tiger snakes are highly flexible predators that track prey resources by means of a complex adaptive response. This response involves not only hard-wired traits that match phenotype to differences in long-term average conditions, but also a developmentally plastic component that matches individuals within each generation to fluctuations in prey size.

Fabien Aubret†‡, Richard Shine§, Xavier Bonnet*§

*CEBC-CNRS, 79360 Villiers en Bois, France

†Université de Poitiers, Conseil Général des Deux-Sèvres, 79360 Villiers en Bois, France

‡School of Animal Biology, University of Western Australia, Crawley, Perth, Western Australia 6009, Australia

§Biological Sciences A08, University of Sydney,

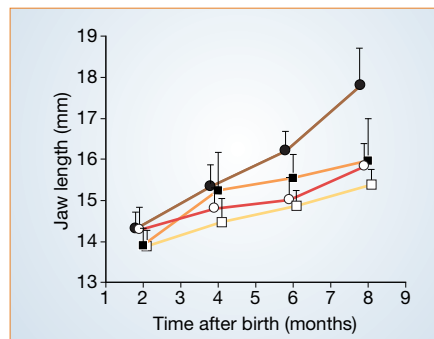


Figure 2 Relative jaw length of growing tiger snakes as a function of prey size. Neonatal snakes from the mainland (Herdsman's Lake; $n = 20$; squares) and from Carnac Island ($n = 19$; circles) were born and kept in captivity. The young snakes were fed weekly on mice. Within each group, half of the snakes were given large prey (mice, 1.8–5.0 g; 2.5–4.7 mm in circumference; filled symbols) and the other half received small prey (mice, 1.7–1.8 g; 2.3–2.5 mm; open symbols). Jaw size of island snakes exposed to large prey increases significantly more than for mainland snakes (mean values $\pm$ s. d.).

New South Wales 2006, Australia

e-mail: rics@bio.usyd.edu.au

1. Bock, W. J. *Am. Zool.* **20**, 217–227 (1980).
2. Dudley, S. A. & Schmitt, J. *Am. Nat.* **147**, 445–465 (1996).
3. Forsman, A. & Shine, R. *Biol. J. Linn. Soc.* **62**, 209–223 (1997).
4. Cogger, H. G. *Reptiles and Amphibians of Australia* (Reed, Chatswood, Australia, 1992).
5. Shine, R. *Herpetologica* **43**, 233–240 (1987).
6. Aubret, F., Maumelat, S., Bonnet, X., Bradshaw, D. & Schwaner, T. *Amphibia–Reptilia* **25**, 9–17 (2003).
7. Pigliucci, M. *Phenotypic Plasticity: Beyond Nature and Nurture* (Johns Hopkins University Press, Baltimore, 2001).

Competing financial interests: declared none.

Ecology

Ultraviolet reflectance by the skin of nestlings

Birds can perceive the reflectance of ultraviolet light by biological structures^{1–5}. Here we show that the skin of the mouth and body of starling nestlings substantially reflects light in the ultraviolet range and that young in which this reflectance is reduced will gain less mass than controls, despite low background levels of ultraviolet and visible light in the nest. We suggest that this ultraviolet reflectance from starling nestlings and its contrast with surrounding surfaces are important for parental decisions about food allocation.

Reflectance of ultraviolet light in the 300–400-nanometre range from feathers, fruits, insects and vole scent marks is used by birds to make decisions about mate choice and food selection^{1–5}. The function of ultraviolet reflectance by the brightly coloured skin of some birds is unknown^{6,7}, but it may play a role in parent–offspring communication⁶.

The flanges (skin surrounding the mouth) in nestlings of both starlings (*Sturnus vulgaris*) and great tits (*Parus major*) is strongly reflective (Fig. 1a,c, respectively; for methods, see supplementary information). In addition, the body skin of starling nestlings from two separate populations (in Switzerland and Britain; see supplementary information) reflects in the ultraviolet (Fig. 1b,e), although the body skin of great-tit (Fig. 1d) and blackbird (*Turdus merula*) nestlings does not⁶.

We investigated the function of this skin reflectance in starlings by applying ultraviolet-light blockers⁸ on the nestlings' bodies and flanges and measuring their body-mass gain after two hours⁹. Nestlings with ultraviolet-reflecting skin ('+UV') gained relatively more mass than non-reflecting ('-UV') nestlings (difference in mass gain $\pm$ s.e.: +UV, 0.23 ± 0.14 g and -UV, -0.24 ± 0.12 g; $F_{1,27} = 4.84$, $P = 0.036$; when included as a random factor in the model (nest^{random}), the nest effect was not significant, $F_{11,27} = 0.30$). In most broods, the mean mass gained by +UV nestlings was greater than that gained by -UV nestlings (sign test, $P = 0.038$). By contrast, we found that

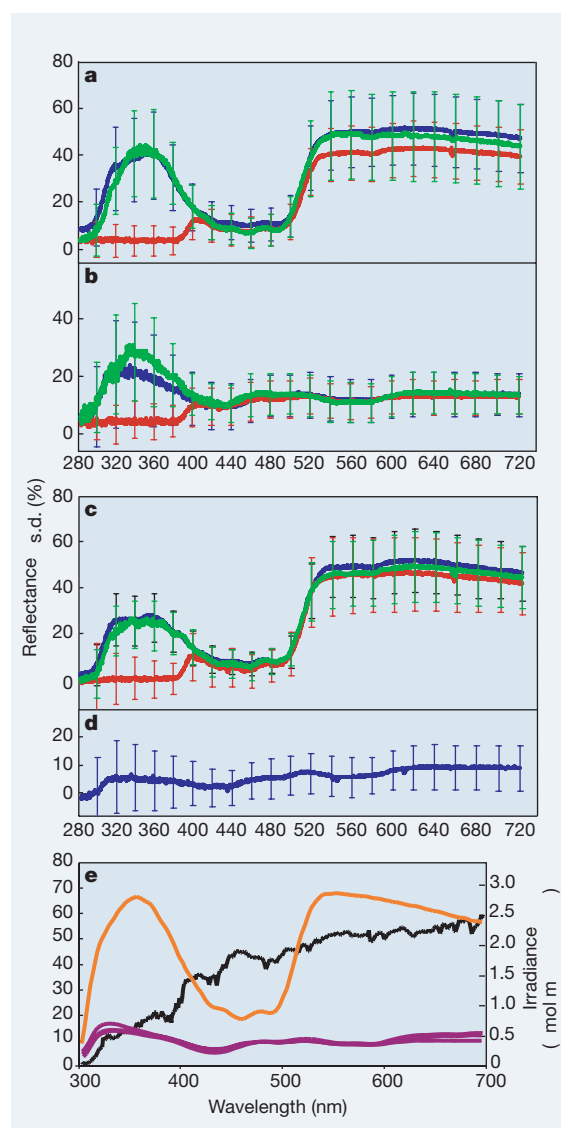


Figure 1 Reflectance spectra of nestling skin and nest irradiance. **a–d**, Reflectance spectra from nestling skin. Starlings: mouth flanges (**a**), $n = 10$; body skin (**b**), $n = 10$. Great tits: mouth flanges (**c**), $n = 25$; body skin (**d**), $n = 22$. Blue, untreated skin; red, skin treated with petroleum jelly containing ultraviolet-light blocker; green, skin treated with petroleum jelly without ultraviolet blocker; curves show medians $\pm$ s.d. **e**, Nest irradiance (black curve; $n = 5$) and reflectance spectra from mouth flanges (orange; $n = 10$) and body skin (purple; $n = 6$) of starling nestlings; median curves are shown.

modifying ultraviolet reflectance on only the mouth flanges in both starling and great tit nestlings did not lead to significant differences in mass gain (starlings: $F_{1,101} = 0.38$, $P = 0.54$, nest^{random} was not significant; great tits: $F_{2,330} = 0.142$, $P = 0.87$, nest^{random} was not significant).

Our results indicate that artificial reduction of the ultraviolet reflectance of body skin and flanges affects mass gain by nestlings. The brightness of skin reflectance in ultraviolet and visible light by starling nestlings did not correlate with either their body mass or tarsus length, but showed a positive relation with their cell-mediated immune responses (see supplementary information; $F_{1,13} = 7.99$, $P = 0.014$; mean brood values: $F_{1,6} = 20.96$, $P = 0.006$). The

brightness of nestling skin reflectance in these birds might therefore signal their phenotypic quality and/or condition^{6,10}.

An alternative explanation could be that nestlings that do not reflect in the ultraviolet might be less easily detected by their parents in dark nests (Fig. 1e) than those that do⁹. We found that ultraviolet-blocking material on the body skin spread after two hours to other young in the brood, reducing the overall body-skin ultraviolet reflectance of the brood. This reduction in ultraviolet contrast between the chicks and their nest, in association with differences in flange ultraviolet reflectance, could lead to less efficient chick detection by the parents, who rely on achromatic or chromatic mechanisms to detect them. This might explain the observed body-mass differences.

The phylogenetic patterns in the evolution of nanostructures⁷ and the pigments responsible for ultraviolet reflectance by skin still need to be determined. Given the likely function of skin reflectance in parent–offspring communication, the ecological factors that affect skin reflectance and the associated parental responses should be investigated.

Violaine Jourdie*†, Benoît Moureau*, Andrew T. D. Bennett‡, Philipp Heeb*

*Department of Ecology & Evolution, University of Lausanne, 1015 Lausanne, Switzerland
e-mail: philipp.heeb@ie-zea.unil.ch

†Department of Zoology, University of Neuchâtel, 2007 Neuchâtel, Switzerland

‡School of Biological Sciences, University of Bristol, Bristol BS8 1UG, UK

1. Cuthill, I. C. et al. *Adv. Study Behav.* **29**, 159–214 (2000).
2. Bennett, A. T. D., Cuthill, I. C., Partridge, J. C. & Maier, E. J. *Nature* **380**, 433–435 (1996).
3. Bennett, A. T. D., Cuthill, I. C. & Norris, K. J. *Am. Nat.* **144**, 848–860 (1994).
4. Sheldon, B. C., Andersson, S., Griffiths, S. C., Örnborg, J. & Sendek, J. *Nature* **402**, 874–877 (1999).
5. Viitala, J., Korpimäki, E., Palokangas, P. & Koivula, M. *Nature* **373**, 425–427 (1995).
6. Hunt, S., Kilner, R. M., Langmore, N. E. & Bennett, A. T. D. *Proc. R. Soc. Lond.* **270**, (suppl.) 25–28 (2003).
7. Prum, R. O. & Torres, R. J. *Exp. Biol.* **206**, 2409–2429 (2003).
8. Andersson, S. & Amundsen, T. *Proc. R. Soc. Lond. B* **264**, 1587–1591 (1997).
9. Heeb, P., Schwander, T. & Faoro, S. *Anim. Behav.* **66**, 637–642 (2003).
10. Kilner, R. M. *Proc. R. Soc. Lond. B* **264**, 963–968 (1997).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Pliocene eclogite exhumation at plate tectonic rates in eastern Papua New Guinea

Suzanne L. Baldwin¹, Brian D. Monteleone¹, Laura E. Webb¹, Paul G. Fitzgerald¹, Marty Grove² & E. June Hill³

¹Syracuse University, Department of Earth Sciences, Syracuse, New York 13244, USA

²University of California at Los Angeles, Department of Earth and Space Sciences, Los Angeles, California 90095, USA

³Commonwealth Scientific and Industrial Research Organization, Petroleum Resources, 26 Dick Perry Avenue, Kensington, Western Australia 6151, Australia

As lithospheric plates are subducted, rocks are metamorphosed under high-pressure and ultrahigh-pressure conditions to produce eclogites and eclogite facies metamorphic rocks. Because chemical equilibrium is rarely fully achieved, eclogites may preserve in their distinctive mineral assemblages and textures a record of the pressures, temperatures and deformation the rock was subjected to during subduction and subsequent exhumation. Radioactive parent–daughter isotopic variations within minerals reveal the timing of these events. Here we present *in situ* zircon U/Pb ion microprobe data that dates the timing of eclogite facies metamorphism in eastern Papua New Guinea at 4.3 ± 0.4 Myr ago, making this the youngest documented eclogite exposed at the Earth's surface. Eclogite exhumation from depths of ~ 75 km was extremely rapid and occurred at plate tectonic rates (cm yr^{-1}). The eclogite was exhumed within a portion of the obliquely convergent Australian–Pacific plate boundary zone, in an extending region located west of the Woodlark basin sea floor spreading centre. Such rapid exhumation ($>1 \text{ cm yr}^{-1}$) of high-pressure and, we infer, ultrahigh-pressure rocks is facilitated by extension within transient plate boundary zones associated with rapid oblique plate convergence.

Oblique collision of major tectonic plates results in rapidly evolving plate boundaries that often exhibit a range of kinematic behaviour along strike as zones of deformation evolve from diffuse boundaries to discrete microplates¹. The evolution of plate boundary zones involves the three-dimensional interplay between both continental and oceanic lithosphere, and mantle asthenosphere over geologic timescales. In Papua New Guinea (PNG), plate boundaries and microplates formed in response to the large-scale oblique collision of the Australian and Pacific plates^{2,3} (Fig. 1). The region preserves a history of arc convergence, continent/arc collision, and ophiolite emplacement⁴ followed by continental rifting and seafloor spreading⁵. In the western Woodlark basin of eastern PNG the transition from seafloor spreading to distributed extension occurs in the most rapidly extending active rift system on Earth^{6,7}. West of the spreading centre, extensional movement on shear zones⁸ led to the formation of metamorphic core complexes in the D'Entrecasteaux Islands^{9,10} and resulted in the rapid exhumation of Pliocene plutonic and metamorphic rocks¹¹, including eclogites interpreted as having formed during prior subduction^{12,13}. The plate boundary zone evolution can be examined in remnants of high-pressure rocks that have escaped overprinting by tectonic or thermal events during subsequent exhumation. We determine the age (4.3 ± 0.4 Myr), and exhumation at plate tectonic rates (tens of kilometres per million years), of the youngest known eclogite exhumed within the obliquely convergent Australian–Pacific (AUS–PAC) plate boundary zone of eastern PNG.

Eastern PNG plate boundary evolution

The present-day tectonic setting of PNG is the result of rapid ($10\text{--}11 \text{ cm yr}^{-1}$) oblique AUS–PAC plate convergence^{2,14–15} (Fig. 1). The AUS–PAC plate boundary zone is complex, has considerable along-strike variation, and has been a location for microplate formation and rotation, short-lived subduction and intraplate deformation^{16,17}. Seismicity and GPS observations define at least three microplates in this convergent plate boundary zone: the Solomon Sea plate, and the North and South Bismark plates. On its northern

boundary the Solomon Sea plate is subducting beneath the South Bismark plate at the New Britain trench. At its southern boundary the tectonic setting is less clear with seismic evidence having been used to argue both for^{14,18} and against¹⁹ southward subduction of the Solomon Sea plate beneath the Trobriand platform at the Trobriand trough. Interaction between the Australian plate and neighbouring microplates define a plate boundary zone presently characterized by arc-continent collision in the Huon peninsula region of PNG (AUS–South Bismark), seafloor spreading in the Woodlark basin (AUS–Woodlark), and possible subduction in the intervening region (Solomon Sea–Woodlark).

The Cenozoic evolution of this complex plate boundary zone involved northwards subduction of the Australian plate beneath the Pacific plate^{4,17} that led to high-pressure metamorphism^{12,13,20} of Jurassic–Cretaceous sediments and basalts, and southward obduction of oceanic crust and mantle of the Pacific plate to form the Papuan ultramafic belt^{4,21} (Fig. 1). In eastern PNG the change from a dominantly convergent plate boundary to a dominantly divergent plate boundary resulted from counterclockwise rotation of the Solomon Sea plate² and ultimately led to seafloor spreading in the Woodlark basin. Since 6 Myr ago the Woodlark basin spreading centre has propagated westwards, separating the once contiguous Woodlark and Pocklington rises⁶.

D'Entrecasteaux Island eclogites

The D'Entrecasteaux Island metamorphic core complexes^{9,10} are located in a zone of extension west of the Woodlark basin spreading centre (Fig. 1). They consist of a lower plate of complexly deformed gneissic domes (up to 2.5 km above sea level), separated from an upper plate of largely undeformed mafic and ultramafic rocks²² by shear zones and detachment faults⁸. The metamorphic basement is divided into two structural zones: a core zone and an outer shear zone^{8,10}. The core zone consists of eclogites, migmatites, gneisses and mylonitic rocks. Although structural fabrics in the core zone are highly variable, an early deformation event (D_1) resulted in gneissic layering that predates peak metamorphism. A second generation of

deformation (D_2) postdates peak metamorphism and resulted in folding and boudinage of the gneissic layering and formation of retrograde tectonic fabrics associated with development of the outer sheared zone. The earliest retrograde fabrics formed during shearing under amphibolite facies conditions at pressures of 7–11 kbar and temperatures of 570–730 °C (ref. 13). Exhumation was facilitated by kilometre-scale mylonitic shear zones whose thermal histories are constrained by $^{40}\text{Ar}/^{39}\text{Ar}$ muscovite, biotite and feldspar, and apatite fission track ages^{11,13}.

Eclogite facies mineral assemblages preserve a high-pressure history and occur as variably retrogressed mafic lenses/layers in quartzofeldspathic gneisses, metamorphosed mafic dikes that cross-cut the D_1 gneissic fabric, and mafic xenoliths in Pliocene granodiorites. Major and trace element abundances are comparable with average MORB values¹². Eclogites record metamorphic conditions that approached 24 kbar and 900 °C (refs 12,13). The eclogite analysed (sample 870921 from Fergusson Island; Fig. 1), preserves garnet ($\text{Alm}_{0.49}\text{Py}_{0.33}\text{Gr}_{0.17}\text{Sp}_{0.01}$; ~30%) + omphacite ($\text{Jd}_{0.57}$; ~40%) along with plagioclase + quartz (~5%) and retrograde amphibole (edenite–pargasite) + biotite (~15%) + chlorite (<5%). Accessory phases include rutile, titanite, apatite and zircon; many of these occur as inclusions within garnet. Omphacite also occurs as inclusions within garnet, as rims on garnet, and as relic grains within edenite–pargasite–plagioclase symplectite. Widespread rutile exsolution in garnet—common in isothermally decompressed high-pressure and ultrahigh-pressure (UHP) rocks²³—occurs in this sample, although not in proximity to zircons analysed, or in the cores of garnet porphyroblasts. Within the matrix, rutile is rimmed by retrograde titanite and is intergrown with ilmenite.

Only omphacite and plagioclase inclusions within the core of a garnet porphyroblast were used for pressure–temperature (P – T) determinations. It is assumed that the omphacite and plagioclase inclusions were in local equilibrium with the encapsulating garnet and that element fractionation reflects ambient P – T conditions. Omphacite inclusions and adjacent garnet compositions indicate peak T conditions^{24,25} of 870–930 °C. Plagioclase and omphacite inclusions within garnet indicate pressures of 20–24 kbar using the jadeite–albite–quartz barometer^{26,27} (Fig. 2).

To determine the age of zircons in eclogite, *in situ* ion probe analyses were performed in thin section, allowing dating of zircons within a textural context with respect to the mineral paragenesis. The majority of the zircons analysed (for example, z2, z3, z4, z15, z26, z37) were relatively small ($\leq 50\text{ }\mu\text{m}$ in longest dimension), elongate to rounded inclusions within garnet porphyroblasts (Fig. 3a). Zircons not enclosed in garnet include the largest zircon analysed (z33), a subhedral $300\text{ }\mu\text{m} \times \sim 70\text{ }\mu\text{m}$ crystal within a retrograde amphibole (edenite–pargasite) and plagioclase ($\text{Ab}_{0.89-0.61}$) symplectite (Fig. 3a). Five analyses were obtained from different portions of this zircon (z33–1–5), with repeated drilling into some areas. Another zircon (z34) occurs within an amphibole rim on a garnet porphyroblast.

The age significance of U/Pb results (Fig. 3b) was determined by assuming that measured Pb was a simple mixture between a single concordant radiogenic endmember and common Pb. The zircons analysed gave a $^{238}\text{U}/^{206}\text{Pb}$ age of $4.33 \pm 0.36\text{ Myr}$ (2σ ; $\text{MSWD} = 3.3$) defined by the intersection with concordia of the best-fit line passing through the data. Because U-series disequilibrium would not shift the results by an amount in excess of the indicated errors²⁸, no corrections were necessary.

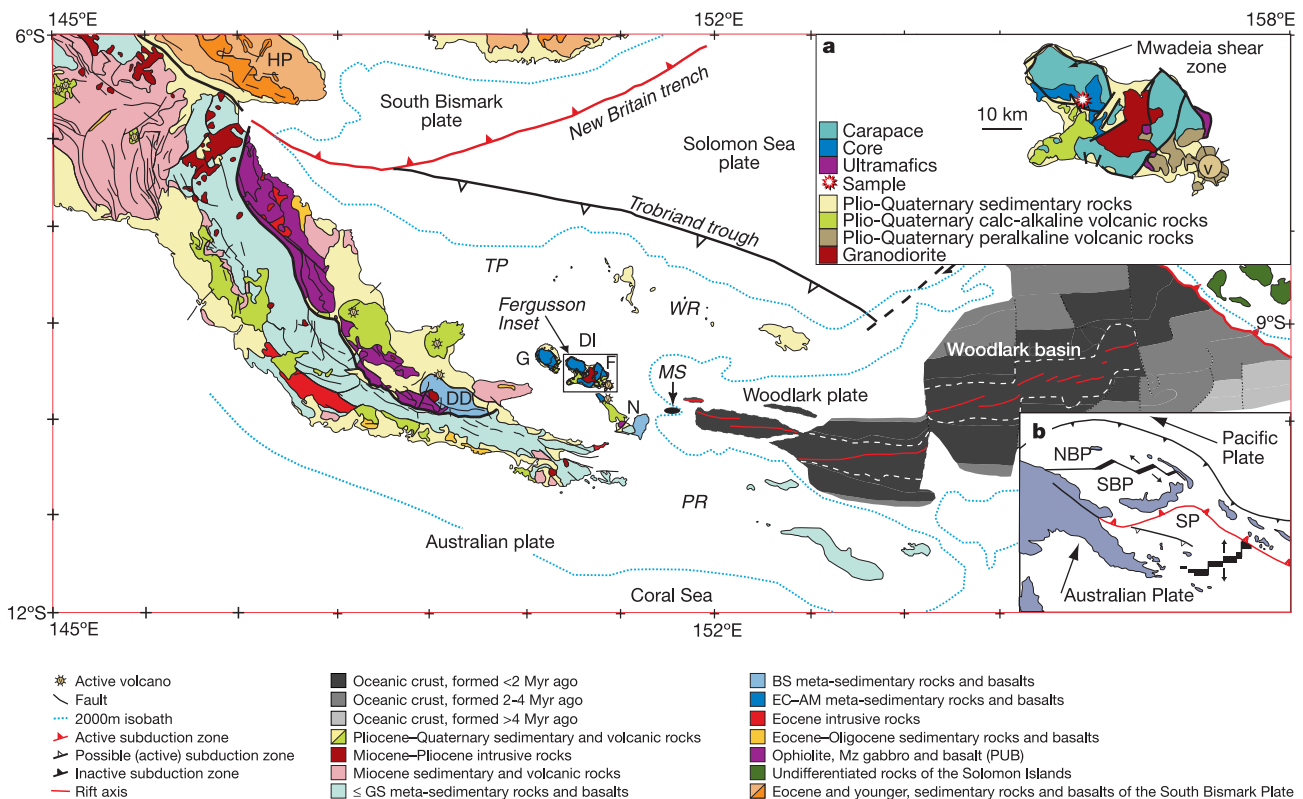


Figure 1 Eastern PNG tectonic and geologic maps^{4,6,21}. DD, Dayman dome; DI, D'Entrecasteaux Islands; G, Goodenough Island; F, Fergusson Island; N, Normanby Island; MS, Moresby seamount; TP, Trobriand platform; WR, Woodlark rise; PR, Pocklington rise; HP, Huon peninsula. PUB, Papuan ultramafic belt; OC, oceanic crust; GS, greenschist; BS, blueschist; EC, eclogite; AM, amphibolite. **a**, Fergusson Island geology (see box on

main figure labelled 'Fergusson Inset'; eclogite locality indicated¹⁰. Carapace, shear zones; core, gneiss, migmatite and eclogite. **b**, Australian–Pacific plate boundary zone microplates². White dashed line indicates Brunhes chron. NBP, North Bismark plate; SBP, South Bismark plate; SP, Solomon plate.

Interpreting U/Pb ages with respect to the eclogite's pressure–temperature–time–deformation (P – T – t – D) history requires establishing whether the zircon was inherited from the protolith, formed during subduction, or formed during exhumation. Lead diffusion in crystalline zircon is too slow to account for complete isotopic resetting of zircon inherited from a protolith²⁹. Thus, given the relatively zirconium-poor bulk composition of the mafic protolith and the lack of identified zircon inheritance, zircon growth most probably occurred during metamorphic recrystallization. Neocrystallization of zircon during metamorphism at P – T conditions lower than upper amphibolite and granulite grade is rare³⁰. Textural evidence indicates that most of the zircons were included within garnets (for example, z2, z3, z15, z26, z37). This textural relationship also argues against an interpretation of melt-controlled or fluid-controlled zircon growth. Other zircons occur in retrograde garnet rims (z34) and within symplectite.

To further constrain the P – T conditions under which the zircons formed, the trace element and rare earth element (REE) geochemistry of zircon and garnet were investigated by *in situ* ion microprobe analysis. Figure 3c shows chondrite-normalized³¹ REE plots for zircon (z33) and adjacent garnet (grt33). Zircon exhibits a positive Ce anomaly relative to chondrite, an overall flat heavy REE pattern, and no significant negative Eu anomaly. Chondrite-normalized patterns for coexisting garnet do not reveal Ce or Eu anomalies. They have steeply positive profiles for Ce to Sm, and negatively sloping profiles for the heavy REE. Zircon/garnet distribution coefficients for the medium to heavy REE increase from 0.66 for Sm to 7.4 for Lu. The form of these REE patterns and the distribution coefficients are consistent with contemporaneous

zircon/garnet growth and are remarkably similar to those determined for alpine eclogites^{32,33}.

The overall similarity to alpine eclogites and the lack of an Eu anomaly provide convincing evidence for zircon growth under eclogite facies conditions³³. Retrograde formation of symplectic intergrowths after garnet + clinopyroxene involved significant plagioclase growth and would have induced a negative Eu anomaly in retrograde zircon. Hence we are confident in interpreting the $^{238}\text{U}/^{206}\text{Pb}$ age of 4.33 ± 0.36 Myr (2σ ; MSWD = 3.3) as marking the time of zircon growth at eclogite facies conditions.

Exhumation rates

Since the Early Pliocene epoch, exhumation of this eclogite from a depth of approximately 75 km was facilitated by tectonic (that is, kilometre-scale mylonitic shear zones⁸) and erosional processes. We can use the boundary condition of the present-day surface of the Earth from which the sample was collected, the U/Pb zircon age and thermobarometric constraints to calculate average exhumation rates for this eclogite. This calculation does not rely on assumptions

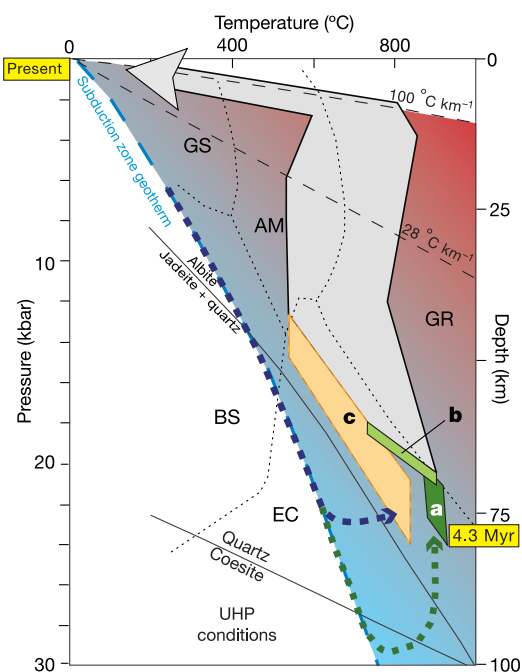


Figure 2 Pressure–temperature conditions for D'Entrecasteaux eclogites. **a**, This study; **b**, from ref. 10; **c**, from ref. 12. Exhumation path based on compilation of P – T data for eclogites, shear zone gneisses, and retrogressed core zone gneisses¹³. Two possible prograde paths for the eclogite are indicated as dashed lines with arrows. Metamorphic facies (abbreviations as in Fig. 1; GR, granulite), coesite = quartz, and jadeite + quartz = albite reaction curves indicated. Mineral compositions are given in the Supplementary Information. Subduction zone geotherm, and present-day range of thermal gradients from the Moresby seamount region⁴³, east of the D'Entrecasteaux Islands, are also shown.

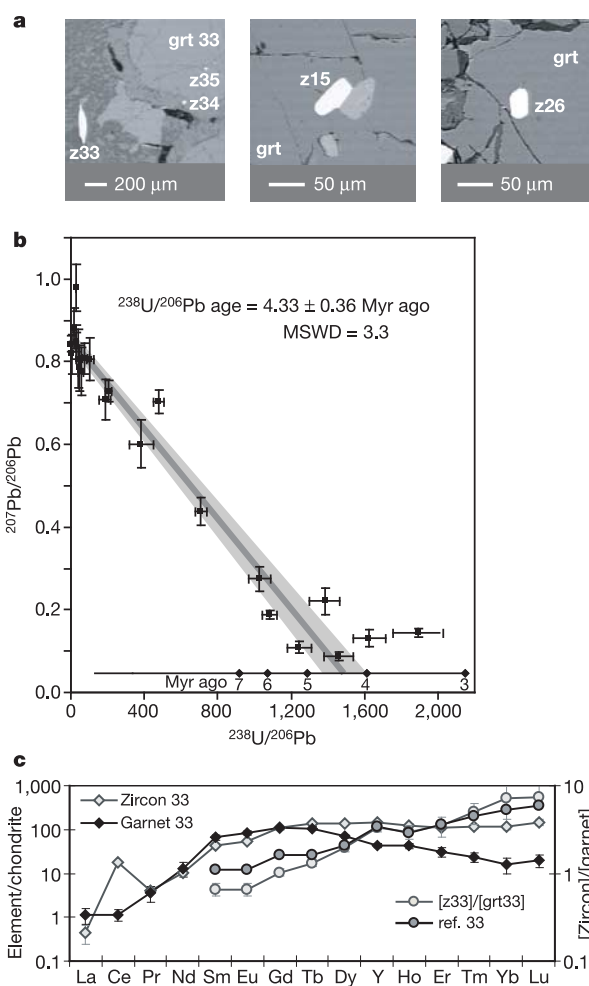


Figure 3 U/Pb, trace and REE results. **a**, Backscattered-electron images of selected zircons. Grt, garnet; z, zircon. **b**, Tera–Wasserburg plot with *in situ* U/Pb ion probe results. Individual analyses are plotted with 1σ errors. Regression (solid line) of U/Pb data, uncorrected for common Pb, indicates a $^{238}\text{U}/^{206}\text{Pb}$ age of 4.33 ± 0.36 Myr ago (2σ ; MSWD = 3.3). Also shown is a 2σ error envelope on the regression. **c**, Trace element and REE data (average $\pm 1\sigma$) for zircon (z33) and garnet (grt33), plotted relative to chondritic values³¹. Also shown are zircon/garnet distribution coefficients³³. U–Th–Pb isotopic and REE data are provided in the Supplementary Information.

related to closure temperatures, or geothermal gradients. A minimum exhumation rate ($\sim 17 \text{ mm yr}^{-1}$) is calculated assuming vertical exhumation. A higher rate of $35\text{--}51 \text{ mm yr}^{-1}$ is calculated given that the eclogite was at $\sim 75 \text{ km}$ at 4.3 Myr ago, was exhumed from beneath the $20^\circ\text{--}30^\circ$ NNE dipping Mwadeia shear zone⁸ and was collected from the Earth's surface (Figs 1 and 2). These average exhumation rates do not consider possible complicating effects such as trench migration and changes in the dip of the shear zone and/or subducting slab. Nor do these rates consider the extent to which mid-crustal flow contributed to changes in crustal thickness during the Australian–Solomon Sea/Woodlark plate boundary zone evolution, addition of mantle derived magmas to the base of the crust, variable exhumation mechanisms, or transient accelerations in exhumation rate. These average exhumation rates are, however, comparable to extension rates near the rift tip ($25\text{--}40 \text{ mm yr}^{-1}$) (ref. 7), half-spreading rates determined both from magnetic reversals (20 and 26 mm yr^{-1}) (ref. 2) and GPS crustal motion surveys ($12\text{--}40 \text{ mm yr}^{-1}$) (ref. 15) in the D'Entrecasteaux–Woodlark region.

At present we cannot constrain the $P\text{--}T\text{--}t$ path followed by this sample before 4.3 Myr ago. If exhumed from UHP depths, 4.3 Myr ago marks the timing of retrograde high-pressure metamorphism (Fig. 2). Alternatively the sample may have reached peak pressure at 4.3 Myr ago and then was subsequently exhumed. Regardless of the details of the prograde path followed, the data presented provide compelling evidence for the existence of eclogite at depths of $\sim 75 \text{ km}$ at 4.3 Myr ago. Movement on major crustal extensional shear zones facilitated exhumation of previously subducted, but largely buoyant crustal material, from beneath mafic and ultramafic rocks³⁴ during rifting. Mineral assemblages preserved in the gneisses of the lower plates record stages of partial re-equilibration on ascent towards the surface¹³. $P\text{--}T$ paths suggest nearly adiabatic decompression at high temperatures, so residence time outside the eclogite stability field must have been short in order to escape total obliteration of high-pressure minerals. Encapsulation of high-pressure minerals and zircon inclusions in strong, fluid impermeable garnet hosts aided in their preservation³⁵. In eastern PNG, preservation of relics of high-pressure mineral assemblages in retrogressed blueschists in the Dayman dome²⁰ and Normanby Island³⁶ also requires rapid exhumation at plate tectonic rates, especially since the present-day Mohorovičić discontinuity (Moho) is elevated relative to crust north and south of the core complexes³⁷. Some of the deeply subducted material may have retraced its path by moving rapidly back up the subduction channel^{4,35} at the same time that the spreading tip was propagating westwards. We infer that rifting reactivated structures within a former plate boundary that previously separated sediments and basalts of largely Australian affinity from overthrust ophiolitic rocks of the Papuan ultramafic belt.

Implications for eclogite exhumation

The D'Entrecasteaux high-pressure rocks are transient crustal features presently exposed at the Earth's surface whose $P\text{--}T\text{--}t\text{--}D$ histories reveal insights into the four-dimensional plate boundary evolution during the transition from subduction to rifting in eastern PNG. This transition was accompanied by an increase in the geotherm (that is, non-steady-state geothermal gradients; Fig. 2), as the plate boundary was reorganized during westward propagation of the Woodlark spreading centre into eastern PNG. As the spreading centre continues to propagate westwards, and crust continues to be thinned, the D'Entrecasteaux Islands would be expected to eventually subside and be thermally overprinted at low pressures owing to impingement of upwelling mantle asthenosphere. This geologic transience may in part explain why similar metamorphic core complexes are difficult to recognize in the geologic record of plate boundary zones.

In the obliquely convergent AUS–PAC plate boundary zone, the

possible role of microplate rotation in eclogite exhumation at plate tectonic rates has important implications for models of exhumation of high-pressure and UHP rocks in other regions. In most high-pressure and UHP terranes geologists are left to infer what structural, petrologic, and thermochronologic data represent in terms of exhumation mechanisms and large-scale plate interactions. Models for the exhumation of UHP rocks generally call on the buoyancy of continental crust subducted to mantle depths as the main driving force for exhumation³⁸, which ultimately manifests itself as regurgitated buoyant crustal slices bound above and below by normal and reverse faults, respectively³⁵. Although satisfactory when considered in a two-dimensional cross-sectional view, these models are difficult to reconcile in three dimensions without the presence of major transfer faults or invoking some element of rotation. An artefact of the buoyancy model is that the rocks become neutrally buoyant in the middle crust, thus requiring an additional mechanism to facilitate exhumation to the surface. In the case of the eastern PNG high-pressure rocks, a complex interaction of microplates within the larger AUS–PAC plate boundary zone plays an important role in their exhumation. Extension in the D'Entrecasteaux Islands may be kinematically linked to microplate rotation and the westward propagation of an oceanic spreading centre into previously subducted crust. For Paleozoic or older high-pressure rocks³⁹, $P\text{--}T\text{--}t\text{--}D$ evidence for the plate boundary zone evolution may have long been obliterated and the transition between subduction and rifting may occur so rapidly that evidence for the timing of these events would potentially be obscured by analytical errors associated with isotopic ages. We note, however, that the average exhumation rate for this eclogite from Fergusson Island is comparable to those obtained from exhumed high-pressure and UHP rocks from other localities (such as Dora Maira and Kokchetav)^{40,41}. Furthermore, it is comparable to exhumation rates of mid-crustal rocks from other portions of the obliquely convergent AUS–PAC plate boundary (for example, the Southern Alps of New Zealand)⁴².

The $^{238}\text{U}/^{206}\text{Pb}$ zircon age of $4.33 \pm 0.36 \text{ Myr}$ indicates that the exhumed lower plate of the D'Entrecasteaux metamorphic core complexes contains the youngest eclogite known on Earth. It formed and was exhumed during the complex reorganization of the Australian–South Bismark–Solomon–Woodlark plate boundary zone of eastern PNG. Results indicate that eclogites can be exhumed at rates that are similar to those at which oceanic crust is created and/or consumed at plate boundaries ($> 1 \text{ cm yr}^{-1}$). □

Methods

Electron probe analysis

Mineral compositions used for thermobarometry were obtained on a $\sim 2\text{-mm}$ diameter garnet porphyroblast and its inclusions by wavelength-dispersive X-ray analysis using a fully automated JEOL 733 electron microprobe at the Rensselaer Polytechnic Institute operating at 15 kV and 21.0 nA . Mineral compositions are given in the Supplementary Information.

U–Th–Pb analysis

For U/Pb ion probe analyses polished thin sections were mounted in epoxy along with zircon standard AS-3, ultrasonically cleaned in 1 N HCl , and Au coated. U–Th–Pb isotopes were measured using the University of California Los Angeles Cameca IMS 1270 ion microprobe. The mass resolution ($5,000$) was sufficient to resolve important molecular interferences in the mass range analysed. During analysis, a 12.5-kV primary O^+ beam was focused on a $\sim 10 \times 15 \mu\text{m}$ area of sectioned zircon from which sputtered positive secondary ions were extracted. The sample surface was flooded with $3 \times 10^{-5} \text{ torr O}_2$ pressure to enhance Pb yields and an aperture was used to block ions derived from the periphery of the ion pit. Analyses consisted of 20 cycles with count times ranging from 1 to 15 s depending on peak intensity for each element.

Trace and REE analysis

Trace and REE data were obtained using the University of California Los Angeles Cameca IMS 1270 ion microprobe. An O^+ primary beam with a current of $10\text{--}11 \text{ nA}$ and an accelerating voltage of $12\text{--}13 \text{ kV}$ was used during analysis. Element to element oxide ratios were determined on doped standard glasses for La, Ce, Nd, Sm, Eu and Gd to correct for isobaric interference on heavier REE isotopes. Zircon and garnet concentrations were derived relative to NIST 610 standard glass concentration, using intensities normalized to ^{30}Si for both zircon and garnet. Beam diameter was focused to $30 \mu\text{m}$ on both grains and

standard glass. Analyses consisted of ten cycles with count times ranging from 0.5 to 10 s depending on peak intensity for each element. Individual concentration values and the standard deviation of the eight analyses for each element are given in the Supplementary Information.

Received 29 March; accepted 12 July 2004; doi:10.1038/nature02846.

1. Stein, S. & Sella, G. F. in *Plate Boundary Zones* (eds Stein, S. & Freymuller, J. T.) 1–26 (American Geophysical Union, Washington DC, 2002).
2. Benes, V., Scott, S. D. & Binns, R. A. Tectonics of rift propagation into a continental margin: Western Woodlark Basin, Papua New Guinea. *J. Geophys. Res.* **99**, 4439–4455 (1994).
3. Hill, K. C. & Hall, R. in *Mesozoic–Cenozoic Evolution of Australia's New Guinea Margin in a West Pacific Context* (eds Hillis, R. R. & Muller, R. D.) (The Geological Society of America, Boulder, CO, 2003).
4. Davies, H. L. Crustal structure and emplacement of ophiolite in southeastern Papua New Guinea. *Coll. Int. CNRS* **272**, 17–33 (1980).
5. Taylor, B., Goodliffe, A. M. & Martinez, F. How continents break up: Insights from Papua New Guinea. *J. Geophys. Res.* **104**, 7497–7512 (1999).
6. Taylor, B., Goodliffe, A. M., Martinez, F. & Hey, R. Continental rifting and initial sea-floor spreading in the Woodlark Basin. *Nature* **374**, 534–537 (1995).
7. Abers, G. A. in *Non-Volcanic Rifting of Volcanic Margins: A Comparison of Evidence from Land and Sea* (eds Wilson, R. C. L., Whitmarsh, R. B., Taylor, B. & Froitzheim, N.) 305–318 (The Geological Society of London, London, 2001).
8. Hill, E. J. Geometry and kinematics of shear zones formed during continental extension in eastern Papua New Guinea. *J. Struct. Geol.* **16**, 1093–1105 (1994).
9. Davies, H. L. & Warren, R. G. Origin of eclogite-bearing, domed, layered metamorphic complexes (“core complexes”) in the D'Entrecasteaux Islands, Papua New Guinea. *Tectonics* **7**, 1–21 (1988).
10. Hill, E. J., Baldwin, S. L. & Lister, G. S. Unroofing of active metamorphic core complexes in the D'Entrecasteaux Islands, Papua New Guinea. *Geology* **20**, 907–910 (1992).
11. Baldwin, S. L., Lister, G. S., Hill, E. J., Foster, D. A. & McDougall, I. Thermochronologic constraints on the tectonic evolution of active metamorphic core complexes, D'Entrecasteaux Islands, Papua New Guinea. *Tectonics* **12**, 611–628 (1993).
12. Davies, H. L. & Warren, R. G. Eclogites of the D'Entrecasteaux Islands. *Contrib. Mineral. Petrol.* **112**, 463–474 (1992).
13. Hill, E. J. & Baldwin, S. L. Exhumation of high-pressure metamorphic rocks during crustal extension in the D'Entrecasteaux region, Papua New Guinea. *J. Metamorph. Geol.* **11**, 261–277 (1993).
14. Pegler, G., Das, S. & Woodhouse, J. H. A seismological study of the eastern New Guinea and western Solomon Sea regions and its tectonic implications. *Geophys. J. Int.* **122**, 961–981 (1995).
15. Tregoning, P. *et al.* Estimation of current plate motions in Papua New Guinea from Global Positioning System observations. *J. Geophys. Res.* **103**, 12181–12203 (1998).
16. Weiler, P. D. & Coe, R. S. Rotations in the actively colliding Finisterre Arc Terrane: paleomagnetic constraints on Plio–Pleistocene evolution of the South Bismark microplate, northeastern Papua New Guinea. *Tectonophysics* **316**, 297–3251 (2000).
17. Hall, R. & Spakman, W. in *Mantle Structure and Tectonic Evolution of the Region North and East of Australia* (eds Hillis, R. R. & Muller, R. D.) 361–381 (The Geological Society of America, Boulder, Colorado, 2003).
18. Cooper, P. & Taylor, B. Seismotectonics of New Guinea: a model for arc reversal following arc-continent collision. *Tectonics* **6**, 53–67 (1987).
19. Abers, G. A. & Roecker, S. W. Deep structure of an arc-continent collision: earthquake relocation and inversion for upper mantle P and S wave velocities beneath Papua New Guinea. *J. Geophys. Res.* **96**, 6379–6401 (1991).
20. Worthing, M. A. Petrology and tectonic setting of blueschist facies metabasites from the Emo metamorphics of Papua New Guinea. *Aust. J. Earth Sci.* **35**, 159–168 (1988).
21. Davies, H. L. & Smith, I. E. Geology of eastern Papua. *Bull. Geol. Soc. Am.* **82**, 8299–8312 (1971).
22. Davies, H. L. & Ives, D. J. The geology of Fergusson and Goodenough Islands, Papua New Guinea. *Aust. Bur. Miner. Resour. Rep.* **82**, 1–83 (1965).
23. Zhang, R. Y. & Liou, J. G. in *Ultra-High Pressure Metamorphism and Geodynamics in Collision-Type Orogenic Belts; Final Report of the Task Group III-6 of the International Lithosphere Project* (eds Ernst, W. G. & Liou, J. G.) 216–228 (International Book Series 4, Bellwether Publishing for the Geological Society of America, Columbia, Maryland, 2000).
24. Ellis, D. J. & Green, D. H. An experimental study of the effect of Ca upon the garnet-clinopyroxene Fe-Mg exchange equilibria. *Contrib. Mineral. Petrol.* **71**, 13–22 (1979).
25. Powell, R. Regression diagnostics and robust regression in geothermometer/geobarometer calibration: the garnet-clinopyroxene geothermometer revisited. *J. Metamorph. Geol.* **3**, 231–243 (1985).
26. Holland, T. J. B. The reaction albite = jadeite + quartz determined experimentally in the range 600–1200°C. *Am. Mineral.* **65**, 129–134 (1980).
27. Gasparik, T. & Lindsley, D. L. in *Pyroxenes: Reviews in Mineralogy* (ed. Prewitt, C. T.) 309–340 (Mineralogical Society of America, Washington DC, 1980).
28. Schmitt, A. K. *et al.* The Geysers-Cobb Mountain-Magma System, California (Part 1): U-Pb zircon ages of volcanic rocks, conditions of zircon crystallization and magma residence times. *Geochim. Cosmochim. Acta* **67**, 3423–3442 (2003).
29. Cherniak, D. J. & Watson, E. B. Pb diffusion in zircon. *Contrib. Mineral. Petrol.* **172**, 198–207 (2000).
30. Hoskin, P. W. O. & Schaltegger, U. in *Zircon* (eds Hancher, J. M. & Hoskin, P. W. O.) 27–62 (Mineralogical Society of America, Washington DC, 2003).
31. McDonough, W. F. & Sun, S.-S. The composition of the Earth. *Chem. Geol.* **120**, 223–253 (1995).
32. Rubatto, D. Zircon trace element geochemistry: partitioning with garnet and the link between U-Pb ages and metamorphism. *Chem. Geol.* **184**, 123–138 (2002).
33. Rubatto, D. & Hermann, J. Zircon formation during fluid circulation in eclogites (Monviso, Western Alps): implications for Zr and Hf budget in subduction zones. *Geochim. Cosmochim. Acta* **67**, 2173–2187 (2003).
34. Martinez, F., Goodliffe, A. M. & Taylor, B. Metamorphic core complex formation by density inversion and lower-crust extrusion. *Nature* **411**, 930–934 (2001).
35. Ernst, W. G. in *Physics of the Earth and Planetary Interiors: Processes and Consequences of Deep Subduction* (eds Rubie, D. C. & van der Hilst, R. D.) 253–276 (Elsevier, Amsterdam, 2001).
36. Baldwin, S. L., Fitzgerald, P. G., Little, T. A., Webb, L. E. & Monteleone, B. D. Exhumation of the youngest high-pressure rocks, at plate tectonic rates, during Plio-Pleistocene continental extension in SE Papua New Guinea. *Geol. Soc. Am. Abstr. Programs* **35**, 556 (2003).
37. Abers, G. A. *et al.* Mantle compensation of active metamorphic core complexes at Woodlark rift in Papua New Guinea. *Nature* **418**, 862–865 (2002).
38. Platt, J. P. Exhumation of high-pressure rocks: a review of concepts and processes. *Terra Nova* **5**, 119–133 (1993).
39. Maruyama, S., Liou, J. G. & Terabayashi, M. Blueschists and eclogites of the world and their exhumation. *Int. Geol. Rev.* **38**, 485–594 (1996).
40. Rubatto, D. & Hermann, J. Exhumation as fast as subduction? *Geology* **29**, 3–6 (2001).
41. Hacker, B. R., Calvert, A. J., Zhang, R. Y., Ernst, W. G. & Liou, J. G. Ultrarapid exhumation of ultrahigh-pressure diamond-bearing metasedimentary rocks of the Kokchetav Massif, Kazakhstan. *Lithos* **70**, 61–75 (2003).
42. Little, T. A., Cox, S., Vry, J. K. & Batt, G. Variations in exhumation level and uplift-rate along the oblique-slip Alpine fault, central Southern Alps, New Zealand. *Geol. Soc. Am. Bull.* (in the press).
43. Shipboard Scientific Party, in *Proc. ODP. Init. Rep. 180* (eds Taylor, B., Huchon, P. & Klaus, A.) 1–77 (Ocean Drilling Program, College Station, Texas, 1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements S.L.B. and P.G.F. gratefully acknowledge support from the US National Science Foundation's Geoscience Directorate, Division of Earth Sciences, Tectonics and Instrumentation and Facilities Programs. M.G. acknowledges support from the Department of Energy. Fieldwork would not have been successful without the efforts of D. Pinasi, and the support from City Resources. A. Schmitt contributed significantly in helping us to obtain ion microprobe REE analyses from zircon and garnet. The ion microprobe facility at UCLA is partly supported by a grant from the Instrumentation and Facilities Program, Division of Earth Sciences, US National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.L.B. (sbaldwin@syr.edu).

The DNA sequence and comparative analysis of human chromosome 5

Jeremy Schmutz¹, Joel Martin², Astrid Terry², Olivier Couronne³, Jane Grimwood¹, Steve Lowry², Laurie A. Gordon^{2,4}, Duncan Scott², Gary Xie^{2,5}, Wayne Huang², Uffe Hellsten², Mary Tran-Gyamfi^{2,4}, Xinwei She⁶, Shyam Prabhakar³, Andrea Aerts², Michael Altherr^{2,5}, Eva Bajorek¹, Stacey Black¹, Elbert Branscomb^{2,4}, Chenier Caoile¹, Jean F. Challacombe⁵, Yee Man Chan¹, Mirian Denys¹, John C. Detter², Julio Escobar¹, Dave Flowers¹, Dea Fotopulos¹, Tijana Glavina², Maria Gomez¹, Eidelyn Gonzales¹, David Goodstein², Igor Grigoriev², Matthew Groza⁴, Nancy Hammon², Trevor Hawkins², Lauren Haydu¹, Sanjay Israni², Jamie Jett², Kristen Kadner², Heather Kimball², Arthur Kobayashi^{2,4}, Frederick Lopez¹, Yunian Lou², Diego Martinez², Catherine Medina¹, Jenna Morgan², Richard Nandkeshwar⁴, James P. Noonan⁷, Sam Pitluck², Martin Pollard², Paul Predki², James Priest³, Lucia Ramirez¹, James Retterer¹, Alex Rodriguez¹, Stephanie Rogers¹, Asaf Salamov², Angelica Salazar¹, Nina Thayer^{2,5}, Hope Tice², Ming Tsai¹, Anna Ustaszewska², Nu Vo¹, Jeremy Wheeler¹, Kevin Wu¹, Joan Yang¹, Mark Dickson¹, Jan-Fang Cheng³, Evan E. Eichler⁶, Anne Olsen^{2,4}, Len A. Pennacchio^{2,3}, Daniel S. Rokhsar², Paul Richardson², Susan M. Lucas², Richard M. Myers¹ & Edward M. Rubin^{2,3}

¹Stanford Human Genome Center, Department of Genetics, Stanford University School of Medicine, 975 California Ave, Palo Alto, California 94304, USA

²DOE's Joint Genome Institute, 2800 Mitchell Avenue, Walnut Creek, California 94598, USA

³Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, California 94720, USA

⁴Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, California 94550, USA

⁵Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

⁶Department of Genetics, Center for Computational Genomics and Center for Human Genetics, Case Western Reserve University School of Medicine and University Hospitals of Cleveland, Cleveland, Ohio 44106, USA

⁷Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA

Chromosome 5 is one of the largest human chromosomes and contains numerous intrachromosomal duplications, yet it has one of the lowest gene densities. This is partially explained by numerous gene-poor regions that display a remarkable degree of noncoding conservation with non-mammalian vertebrates, suggesting that they are functionally constrained. In total, we compiled 177.7 million base pairs of highly accurate finished sequence containing 923 manually curated protein-coding genes including the protocadherin and interleukin gene families. We also completely sequenced versions of the large chromosome-5-specific internal duplications. These duplications are very recent evolutionary events and probably have a mechanistic role in human physiological variation, as deletions in these regions are the cause of debilitating disorders including spinal muscular atrophy.

The US Department of Energy's interest in chromosome 5 emerged from a series of pilot studies begun at the Lawrence Berkeley National Laboratory focusing on a cluster of interleukin genes located at human 5q31. The insights gained from these detailed analyses of a single megabase of chromosome 5 illustrated how finished human sequence could contribute to gene annotation and how multi-mammalian sequence comparisons could lead to the sequence-based identification of noncoding elements possessing gene regulatory activities^{1–3}. The finished sequence of chromosome 5 and its analysis alone and in comparison to orthologous regions in other vertebrate genomes now provides a chromosome-wide catalogue of genes and evolutionarily conserved noncoding sequences. Many of these observations, as well as clues into disease-causing deletions arising from the segmented duplication landscape of chromosome 5, can only now be appreciated upon finishing the sequence of this chromosome.

Mapping and sequencing

After the completion of the initial draft sequencing in 2001 we selected clones with an approach that integrated all of the public sequence, previously reported clone contigs^{4–6} including the Celera scaffolds⁷, bacterial artificial chromosome (BAC) and fosmid end sequences, and BACs isolated with an overgo hybridization strategy to close gaps between anchored contigs. The final version of the tiling path contains 1,763 clones, (96% BACs) with four gaps remaining, all in the long arm. None of these remaining gaps are part of the large chromosome 5 duplications, and they appear to be unclonable in current vector systems. In addition, our standard strategy of seeding and then walking into gaps based on restriction

maps proved unworkable in the duplication region of 5q13 associated with spinal muscular atrophy (SMA), and led to mapping errors with its primary insertion copy at 5p14 and secondary copy at 5p13. Therefore, we adopted a strategy of drafting high depth clone coverage from the single individual RPCI-11 BAC library in order to construct single haplotype paths spanning the duplications.

On the basis of internal and external quality checks, we estimate the accuracy of our finished sequence to exceed 99.99%⁸. In total, we finished 177,702,766 base pairs (bp) and estimate the total chromosome size, including the clone gaps and the recalcitrant centromeric and subtelomeric regions, to be 180.8 megabases (Mb). The finished

Table 1 **Chromosome 5 sequence features**

Feature	Value
Sequence length (bp)	177,702,766
G+C content (%)	39.5
Gene loci	923*
Known	827
Novel	55
Putative	41
Non-processed pseudogenes	98
Processed pseudogenes	479
tRNAs	20
tRNA pseudogenes	4
Total repeat content (bp)	82,349,155 (46.3%)
Alu	14,998,401 (8.4%)
LINE 1	32,864,033 (18.5%)
LINE 2	4,757,270 (2.7%)
Simple and low complexity	2,594,624 (1.5%)
Other	27,134,827 (15.3%)

*With 1,598 full-length transcripts.

sequence covers 99.9% of the euchromatic sequence and captures all known genes that were previously mapped to chromosome 5 (T. Furey, personal communication). The Stanford v.4 G3 radiation hybrid map⁹ was compared to the sequence and it matched the marker order well (see Supplementary Fig. S1). Thirteen (out of 442) unplaced markers were found to have been originally incorrectly assigned to chromosome 5. Recombination distances from the deCODE¹⁰ meiotic maps were compared to physical distances with recombination rates accurately tracking physical distance (see Supplementary Fig. S2), as previously reported for other chromosomes^{11–13}.

Gene catalogue

We placed gene model transcripts on the chromosome 5 sequence and manually reviewed these models using previously described methods¹¹ (Table 1). Ultimately, 923 protein-coding regions were verified as gene loci (see Supplementary Table S1 and http://www.jgi.doe.gov/human_chr5). These loci contain 1,598 full-length (or nearly full-length) transcripts, including partial evidence for additional splice variants (see Supplementary Information). Loci were placed in three categories: 'known', 'novel' and 'pseudogenes', consistent with our previous definitions¹¹. Transcripts for which a unique open reading frame (ORF) could not be determined and putative genes defined by *ab initio* models but with no supporting experimental evidence were not considered valid. A total of 827 known loci were identified based on 2,203 RefSeq genes and other full-length complementary DNA sequences in GenBank, extending 36% of RefSeq transcripts by more than 50 bp at the 5' end and 18% at the 3' end, while maintaining the original ORF. Gene loci 3' ends were not extended when the only evidence was from rare expressed sequence tag (EST) variants. Evidence for 55 novel loci was supported by full-length cDNA sequence, spliced ESTs, and/or similarity to known human or mouse gene sequences. Forty-one putative gene loci were modelled using orthologous mouse cDNA sequences. Twenty transfer RNA genes and four tRNA pseudogenes were predicted, similar in density to other finished chromosomes^{11–13}.

The extent of alternative splicing was characterized based on the existing cDNA and EST data. Considering only messenger RNA sequences in GenBank, 1,598 distinct transcripts were identified, providing an average coverage of 1.7 annotated transcripts per locus (see Supplementary Information). These mRNAs provide strong evidence for alternative splicing of 408 (44%) of the 923 loci, each having two or more associated transcripts. A total of 577 pseudogenes and pseudogene fragments were also identified, representing two classes: (1) 98 non-processed pseudogenes that display a structure similar to the parent locus and probably resulted from genomic duplication events; (2) 479 processed pseudogenes that

presumably resulted from viral retrotransposition of spliced mRNAs (see Supplementary Information). No significant bias towards over-representation of pseudogenes from a particular gene family was observed.

Chromosome 5 genomic duplications

We performed a detailed analysis of duplicated sequence ($\geq 90\%$ identity and ≥ 1 kilobase (kb) length) by comparing chromosome 5 against the July 2003 human genome assembly. An estimated 3.49% (6.26 Mb) of the chromosome consists of segmental duplications, lower than the genome-wide average of 5.3% (see Supplementary Table S2 and Supplementary Fig. S4). Chromosome 5 segmental duplications, however, show a higher degree of sequence identity ($\geq 97.5\%$), especially with other regions of chromosome 5 (see Supplementary Fig. S5), than do the duplications on other chromosomes. Intrachromosomal duplications are clustered in ten regions (Fig. 1) and represent the majority of the gene duplications, including the largest gene family: the protocadherins (see Supplementary Information). The high degree of sequence identity underlying most of these intrachromosomal genomic duplications suggests that these structures are relatively recent duplications or gene conversion events that emerged during the separation of humans and the great apes (see Supplementary Fig. S3 and Supplementary Table S2).

Subtelomeric and pericentromeric biases have been reported for segmental duplications for other human chromosomes. Despite the fact that large tracts of alpha-satellite DNA have been sequenced on both chromosomal arms near the centromere, there is little evidence for extensive pericentromeric duplication, with 5p11 showing almost a complete absence of duplications. A single duplication in 5q11 (96% identity over 250 kb) between chromosomes 1 and 5 accounts for nearly all pericentromeric duplicated bases. The pericentromeric region of chromosome 5, along with 19q11, may define a duplication-quiescent model of pericentromeric organization. The telomeric regions do show extensive interchromosomal duplications (Fig. 1), with 25% (2.48 out of 9.08 Mb) of all interchromosomal alignments occurring within 2 Mb of the long arm telomeric repeat sequence (see Supplementary Table S3).

SMA duplication region

One of the most duplicated regions on chromosome 5 occurs in a 1–2-Mb interval in 5q13.3. Homozygous deletions of the *SMN1* gene and variable copies of the *SMN2* duplication in this region have been associated with various forms of spinal muscular atrophy and susceptibility to the disease^{14,15}. Analysis of carriers and controls suggests extreme locus variability, but the underlying structural variation has never been documented at the sequence level¹⁶. We identified a complex arrangement of 311 pairwise alignments

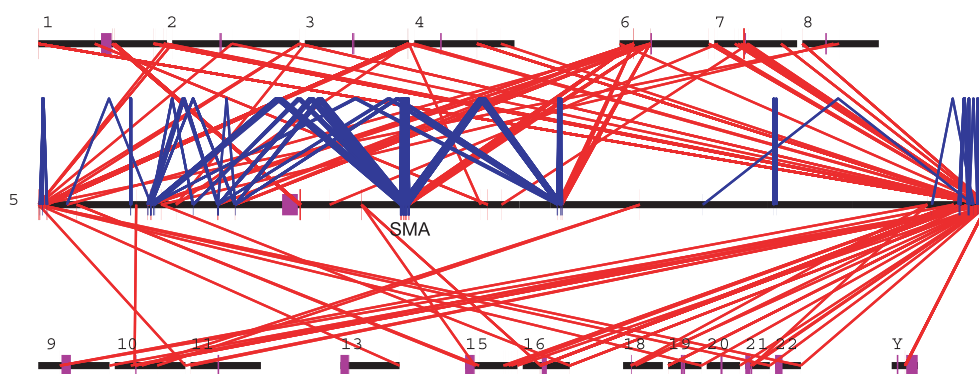


Figure 1 Distribution of segmental duplications on chromosome 5. Large (>5 kb) highly similar ($>90\%$) intrachromosomal (blue) and interchromosomal (red) segmental

duplications are shown for chromosome 5. Chromosome 5 is drawn at a greater scale than the other chromosomes. The centromeres are depicted as purple bars.

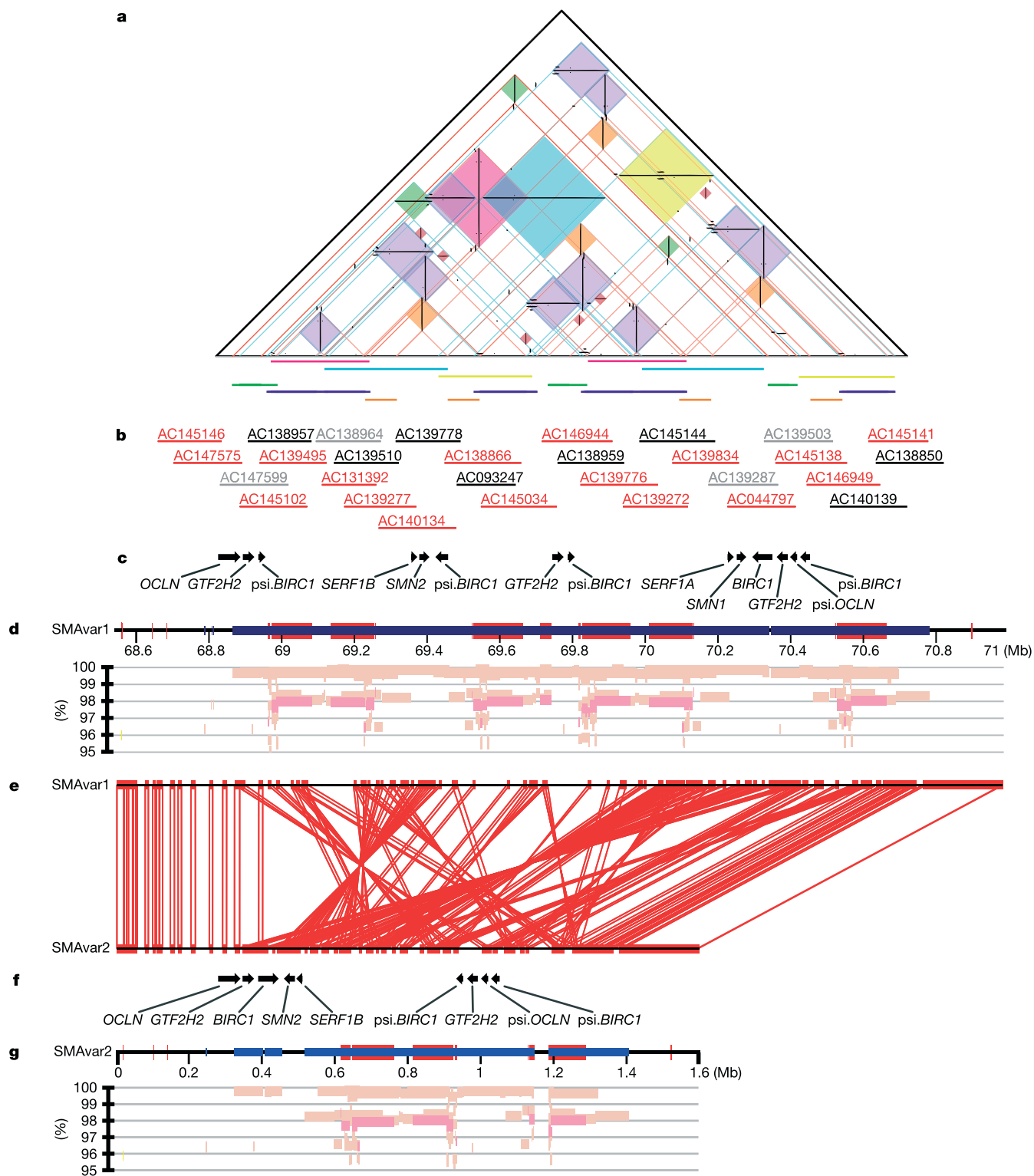


Figure 2 Diagram of the SMA region showing both SMAVar1, the published variant, and SMAVar2, the alternative RPC11 variant. **a**, Self_dot_plot⁴⁴ (http://staffa.wi.mit.edu/page/Y/azfc/self_dot_plot.pl) of SMAVar1. Vertical bars represent inverted repeats, horizontal bars direct repeats. Each dot represents a 200-bp perfect match. The three largest repeats are coloured pink, blue and yellow. **b**, RPCI-11 BAC clone path through SMAVar1. Red clones are in the final path, black clones are finished, grey clones are unfinished. **c**, Gene content of SMAVar1. **d**, The duplication pattern for SMAVar1 is shown along the

scale. Interchromosomal (red) and intrachromosomal duplications (blue) are indicated. The underlying pairwise alignments of segmental duplications (>95% > 1 kb) are depicted as a function of per cent identity (below the horizontal line) with different colours corresponding to the location of the pairwise alignment on different human chromosomes (light pink = 5; dark pink = 6; yellow = 3). **e**, A comparison of the interhaplotype structure between the two variants using Miropeats⁴¹ with a threshold of 7,000. **f**, Gene content of SMAVar2. **g**, Duplication pattern for SMAVar2.

representing the SMA region (Fig. 1). On average, the duplications are long (~200 kb) and show a high degree of identity (98.66%). Duplications in this region include interchromosomal duplications, all of which map to chromosome 6, with three very large tandem

(>99.5% identity) and other various interspersed intrachromosomal duplications (Fig. 2). Interestingly, this region is enriched in genes. We annotated 14 loci in this region, including *SERF1* (small EDRK-rich factor 1), *BIRC1* (baculoviral IAP repeat-containing 1)

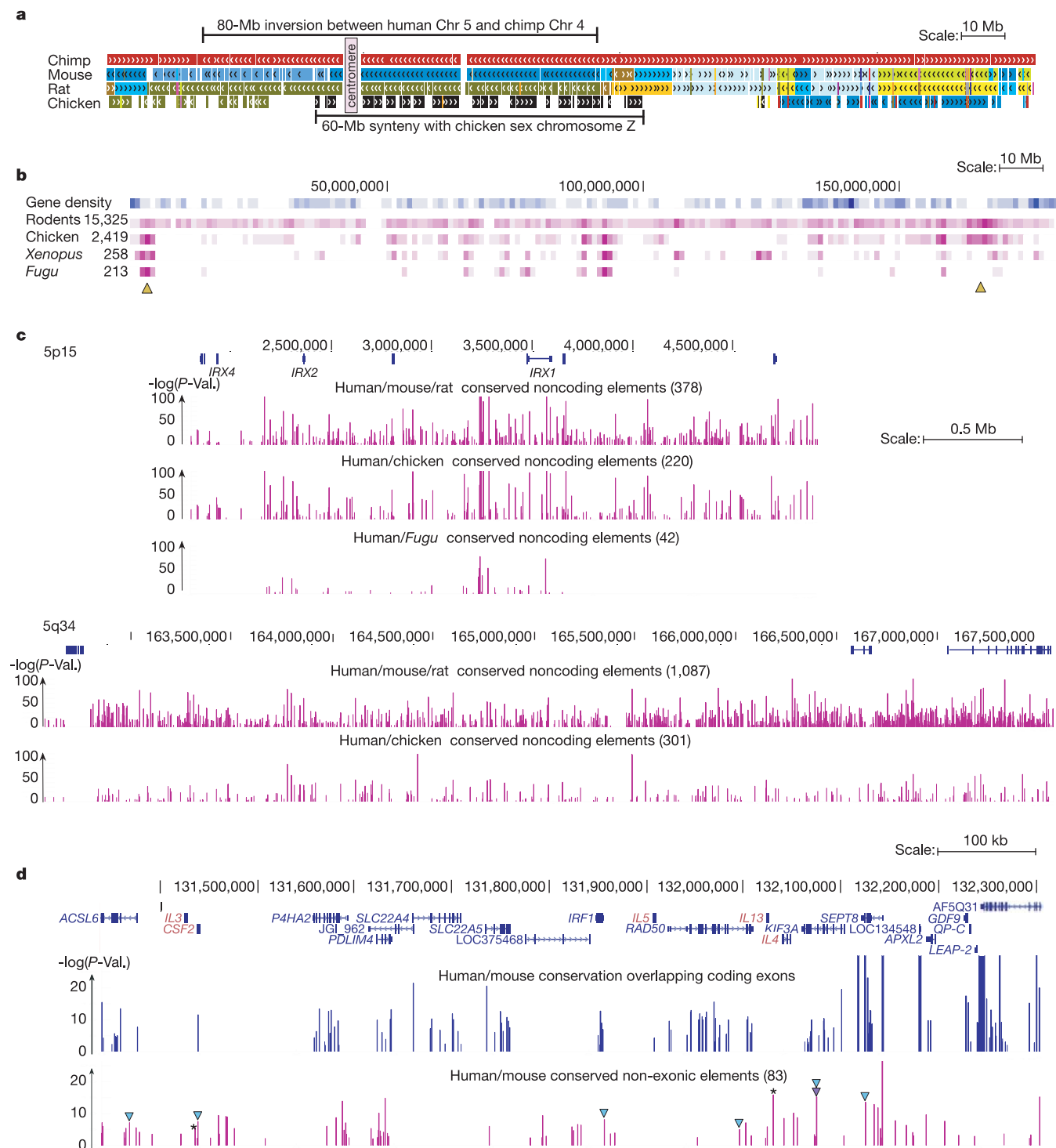


Figure 3 Comparative biology. **a**, Segmental homology maps between human chromosome 5 and the mouse, rat and chicken genomes (see Methods). **b**, Noncoding conservation density. The plot shows the normalized density of the human/mouse/rat, human/mouse/chicken, human/mouse/*Xenopus* and human/mouse/*Fugu* conserved elements. Yellow triangles indicate the location of regions expanded in panels c and d. **c**, The two largest human/mouse/rat/chicken homologous segments overlap gene-poor regions with a high density of conserved noncoding elements (see text). **d**, Interleukin

region. The first plot shows conservation overlapping coding exons, the second plot shows non-exonic conservation. Blue triangles indicate uncharacterized elements conserved in chicken; Purple triangles show uncharacterized elements conserved in *Xenopus*; asterisks are known interleukin enhancers³. These are conserved only in rodents (see text). For clarity only one isoform per gene is shown. In c and d conserved elements are ranked by their statistical significance relative to the local neutral mutation rate. The height of the bars is proportional to $-\log(P\text{-value})$ (PEAK-VISTA; see Methods).

and *SMN* (survival of motor neuron), the gene for SMA.

During the sequencing and assembly of this region, we generated a consensus sequence for a second haplotype variant from the RPCI-11 BAC library. Both haplotypes represent high-quality finished sequence and differ only by a remaining ~50-kb clone gap within SMAvar2. Sequence comparison of these regions (SMAvar1 against SMAvar2) revealed extensive variation. At least two large-scale rearrangements (>400 kb) and multiple smaller insertion/deletion events are required to reconstruct an ancestral haplotype. Although there are many scenarios for the evolution of these variants, one explanation may be that a portion of the SMAvar2 region (0.3–0.9 Mb) was inverted (68.9–69.4 Mb) and subsequently duplicated in SMAvar1 (69.8–70.4 Mb). Such extensive structural variation between haplotypes may not be uncommon in regions of extensive segmental duplication.

Comparative biology

To understand further the evolution and functional sequences of human chromosome 5, we performed comparative analyses against the available chimpanzee, mouse, rat, chicken, frog (*Xenopus tropicalis*) and fish (*Fugu rubripes*) draft genomes. These comparisons revealed numerous large-scale chromosomal rearrangement events occurring since each of these species last common ancestor with humans, as well as a variety of nonrandomly distributed conserved noncoding regions (Fig. 3a). Additional analyses of the distribution of genes and conserved noncoding sequences along the length of the chromosome support the existence of large gene-poor regions with highly conserved noncoding sequences that may regulate genes from a distance. Furthermore, we examined conservation in a comparative analysis of the extensively studied interleukin gene cluster.

Synteny

By building segmental maps from DNA alignments of all the vertebrate species described above, we were able to confirm and extend previous homologous chromosomal relationships with human chromosome 5. Whereas recent experimental studies support that large-scale rearrangements (40–175 kb) have frequently occurred during primate genome evolution¹⁷, our comparison of chromosome 5 and the recent chimpanzee draft genome sequence (International Chimpanzee Genome Sequencing Consortium, manuscript in preparation) uncovered even larger-scale events. For example, we found a large 80-Mb inversion in comparison to the chimpanzee genome, homologous to almost half of human chromosome 5 between 5p14 and 5q15 (Fig. 3a). This finding using the genomic draft data independently confirms previous fluorescence *in situ* hybridization (FISH) experiments¹⁸. It has been proposed that these large-scale rearrangements create barriers to fertile mating and triggered the speciation that separated these two lineages¹⁹. Comparison with the mouse genome sequence²⁰ yielded 142 chromosomal rearrangements ranging in size from 200 kb to 17 Mb. Between human and chicken, we found that one-third of chromosome 5 is homologous to the chicken sex chromosome Z²¹, further indicating that sex chromosomes have evolved independently after the avian and mammalian split some 300 million years ago²².

Chimpanzee

In addition to exploring the syntenic relationship between chromosome 5 and the chimpanzee draft assembly, we catalogued sequence changes between these two primates. To explore the constraint on human–chimpanzee evolution in noncoding regions, we compared the number of nucleotide substitutions in coding sequences, as well as noncoding regions conserved and not conserved in rodents. We found a substitution rate of 0.0067 changes per nucleotide in coding sequences, 0.0091 in noncoding regions conserved in rodents, and 0.015 in noncoding regions not conserved in rodents. The decreased

substitution rate in coding sequences and noncoding sequences conserved in rodents (compared to noncoding regions not conserved in rodents) support the theory that both of the former categories are under evolutionary constraint. This also supports the theory that human/chimpanzee coding and noncoding sequences conserved in rodents have been under moderate selective constraint since the last common human/chimpanzee ancestor. We next compared the patterns of variation within human and chimpanzee exons to identify genes potentially under positive selection in the human lineage as reported in ref. 23. We found 21 genes randomly distributed over chromosome 5 displaying a *P*-value less than 0.01 for an increased evolutionary rate in humans. Of note is that the two highest ranked genes (*FBN2* and *SQSTM1*) are both linked to human diseases. Mutations in *FBN2* cause pathologies similar to Marfan syndrome (*FBN1*), whereas *SQSTM1* has been linked to Paget's disease of the bone²⁴. As the chimpanzee genome reaches a further draft state, a similar complete re-analysis of the entire human gene set will probably yield large numbers of quickly evolving genes, which may explain unique aspects of human biology.

Vertebrate conservation

To annotate functional elements, we identified slowly evolving regions, presumably under evolutionary constraint^{25,26}, through DNA comparison with rodent, chicken, *Xenopus* and *Fugu* (*P*-value <0.01). A chromosome-wide analysis resulted in 15,325 discrete noncoding regions conserved between human/mouse/rat, 2,429 between human/mouse/chicken, 258 between human/mouse/*Xenopus* and 213 between human/mouse/*Fugu*. We found that the distribution of human/mouse/*Fugu* conserved noncoding sequences was highly uneven along the chromosome (Fig. 3b), with 42 centred in 5p15 around an Iroquois homeobox (*IRX*) gene family. These discrete evolutionarily conserved sequences represent a prioritized substrate for future experimental studies to elucidate their function and potential role in gene regulation.

Gene-poor regions

Recent work has shown that a significant fraction of noncoding elements conserved between human and *Fugu* has gene regulatory activity even though many are located at great distances from the genes whose expression they control²⁷. In addition to their location between conserved flanking genes, evidence to support distant gene regulatory sequences is found in the maintenance of long syntenic blocks across distant evolutionary species²⁸. To determine whether such regions exist on human chromosome 5, we built a segmental homology map between human, chimp, mouse, rat and chicken. This map revealed two segments larger than 3 Mb that do not contain any evolutionary break points or insertions (>250 kb) within all examined species. Notably, despite this high level of conservation, these two large segments have very few known genes and overlap the extreme gene-poor regions at 5p15 (3.1 Mb) and 5q34 (5.0 Mb). In addition, each is highly enriched for conserved noncoding sequences with distantly related non-mammalian vertebrates (Fig. 3c). In contrast to the interleukin cluster (described below) and despite being gene poor, the 5p15 region contains 378, 220 and 42 noncoding elements conserved in rodents, chicken and *Fugu*, respectively³. A similar level of noncoding conservation was observed in the 5q34 gene desert region containing 1,087 noncoding elements conserved with rodents, 301 with chicken, but none with *Fugu*. Although functional studies are needed to determine whether these ancient conserved sequences regulate the limited number of genes in these regions, it is interesting to note that the 5p15 region contains a cluster of *IRX* genes that have multiple roles during pattern formation in vertebrate development. The high density of conserved noncoding elements with extended synteny in these gene-poor regions suggests that these regions contain elements that regulate distant genes.

Interleukin cluster

The interleukin gene cluster on 5q31 is a region of particular interest to immunologists because of the presence of five haematopoietic growth factor genes (*IL3*, *CSF2*, *IL5*, *IL13* and *IL4*) and two quantitative trait loci associated with atopic asthma and Crohn's disease susceptibility. From the comparative analysis of this 1 Mb of sequence, we found that 140 of the 190 (76%) human coding exons overlap regions conserved in mouse. This number decreased to 126 (66%) when examining human/mouse/chicken conservation (P -value <0.01 ; Fig. 3d; see also Supplementary Table S4). Consistent with the known fast evolutionary rate of the interleukin genes, most of the interleukin exons (18 of 21) are among the exon sequences that lack similarity between the species. In the analysis of noncoding sequences, we found 83 conserved human/mouse elements that include two previously characterized gene enhancers (CNS-1 and CNS-7)²². One of these elements is more highly conserved than CNS-1 and CNS-7, yet remains functionally undefined. In addition, we found six human/mouse/chicken conserved noncoding sequences, one of which is also conserved in *Xenopus*.

Human disease

Not long after the concept of using anonymous polymorphic DNA markers to localize disease loci was proposed, linkage to many diseases on chromosome 5 was found, and positional cloning and other strategies rapidly isolated the genes for these clearly segregating disorders. So far, mutations in 66 specific genes are known for mendelian diseases (see Supplementary Table S5); an additional 14 single-gene diseases have been mapped to chromosome 5 but have not yet been linked to specific genes. In one of the first examples of a study taking advantage of linkage disequilibrium to positionally clone a gene, ref. 29 identified the DTD gene mutated in diastrophic dysplasia in the Finnish population in 1994. Identification of mutations in the growth hormone receptor gene, at 5p12-p13, in Laron dwarfism was an early case of 'positional candidate cloning', in which the gene was cloned and its location known before mapping the trait³⁰. In addition to SMA, microdeletions in a duplicated region in 5q35 cause Sotos syndrome, a debilitating disorder that results in cranial overgrowth and mental retardation³¹, in which the duplication is thought to mediate severity³². The availability of this completed sequence will further advance our understanding of human disease, and the rate at which disease genes are identified and cloned with causative mutations should be greatly accelerated. □

Methods

Mapping and sequencing

We seeded chromosome 5 with P1, PAC and Caltech BAC clones anchored to a set of 1,645 radiation hybrid markers and known genes, mapping 5,392 clones to chromosome 5 and with 4,943 of these localized by FISH. After constructing a single enzyme restriction digest map, we chose a minimal tiling path. For the SMA duplication regions, hybridization probes were designed at 50-kb intervals across the working maps with additional probes for each uniquely identified duplcon and screened against RPCI-11. Results were binned and ~40% of positives selected for sequencing. Single haplotype maps were constructed by sequence analysis, relying on >30-kb alignments with zero or one discrepancy and multiple clone depth. For the complex 5q13 copy, we used an iterative cycle of probing, sequencing, direct repeat resolution, finishing and re-analysis.

We generated sequence by using a clone-by-clone shotgun sequencing strategy³³ followed by finishing with a custom primer approach. BAC DNA was sheared by using a Hydroshear Instrument (GeneMachines), size selected (3–4 kb) and subcloned into the vector pUC18. Randomly selected subclones were sequenced in both directions using universal primers and BigDye Terminator chemistry to an average depth of $\times 8$. Sequences were assembled and edited by using the Phred/Phrap/Consed suite of programs^{34,35}. After manual inspection of the assembled sequences, clones were finished by re-sequencing and by sequencing off of plasmid subclones or the large insert clone by using custom primers. All finishing reactions were performed with dGTP BigDye Terminator chemistry (Applied Biosystems). Clones with high repeat content or that showed considerable bias when cloned into pUC18 had additional 8–10-kb libraries constructed in a low copy number vector. Recalcitrant areas and difficult to sequence gaps were closed with sequence data derived from transposon sequencing, small insert shatter

libraries³⁶, or PCR. Each clone was finished according to the agreed international standard for the human genome (<http://genome.wustl.edu/Overview/g16stand.php>).

Marker placement

Genetic markers were placed on the genomic sequence using electronic PCR³⁷. Markers were allowed to have up to three mismatches and were subsequently verified by placing the STS sequence (downloaded from UniSTS) via NCBI Megablast using a drop-off value of 180, a match reward of 10, a gap penalty of -20 , and a word size of 22.

Pseudogene identification

Pseudogenes were defined as gene models built by homology to known human genes where alignment between the model and the homologue shows at least one stop codon or frameshift mutation. For the fragments of chromosome 5 genomic sequence that were masked of repeats by using RepeatMasker (A. Smit and P. Green, unpublished data)³⁸, we identified homology to human IPI proteins by using NCBI BLASTX. For each fragment of genomic sequence homologous to an IPI protein, we built gene models by using the GeneWise program. The overlapping gene models were clustered and the alignment of the top-scoring model with its human homologue was analysed for the presence of stop codons and frameshifts. The models were then manually analysed to confirm pseudogene status. Sequences of 431 processed pseudogenes that had been identified previously³⁹ were mapped to the genomic sequence of chromosome 5 by using the BLAT tool. Loci with multi-exon mapping, overlaps with the pseudogenes described above, and simple repeats identified by RepeatMasker were eliminated. Pseudogene status of the remaining sequences was manually validated.

Segmental duplication analysis

We used a BLAST-based detection scheme⁴⁰ to identify all pairwise similarities representing duplicated regions (≥ 1 kb and $\geq 90\%$ identity) within the finished sequence of chromosome 5 and compared to all other chromosomes in the NCBI genome assembly (build 34). A total of 1,818 pairwise alignments representing 16.57 Mb of aligned base pairs and 6.26 Mb of non-redundant duplicated bases were analysed on chromosome 5. The program Parasight (J. A. Bailey, unpublished data) was used to generate images of pairwise alignments. We also analysed pairwise alignments for per cent identity and the number of aligned bases. Satellite repeats were detected by using RepeatMasker (version 15 May 2002) on slow settings. Analysis of haplotype structural variation was performed using the program Miropeats (threshold = 7,000)⁴¹.

Comparative analysis

In this work, we used the following genomic assembly builds: chimpanzee November 2003, mouse October 2003, rat June 2003, chicken February 2004 (from <http://genome.ucsc.edu/>), *X. tropicalis* v1.0 and *F. rubripes* v3.0 (from <http://jgi.doe.gov/>). All the segmental homology maps in n -dimensions are computed using PARAGON (v2.13; O. Couronne, unpublished data). As input for PARAGON, we used BLASTZ (v6)⁴² DNA pairwise alignments of all the species to human. Slowly evolving regions are extracted from the alignments using PEAK-VISTA (P -value >0.01 ; S. Prabhakar, unpublished data). We built a four-dimension human/chimp/mouse/rat segmental homology map with PARAGON, aligned all the segments with MLAGAN (v12)⁴³ and computed the slowly evolving conserved regions with PEAK-VISTA. Interleukin homology among species was extracted from the PARAGON segmental map, built with MLAGAN multiple alignments; the slowly evolving conserved regions were extracted with RANK-VISTA.

Received 10 May; accepted 2 August 2004; doi:10.1038/nature02919.

1. Frazer, K. A. *et al.* Computational and biological analysis of 680 kb of DNA sequence from the human 5q31 cytokine gene cluster region. *Genome Res.* **7**, 495–512 (1997).
2. Symula, D. J. *et al.* Functional screening of an asthma QTL in YAC transgenic mice. *Nature Genet.* **23**, 241–244 (1999).
3. Loots, G. G. *et al.* Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science* **288**, 136–140 (2000).
4. Church, D. M., Yang, J., Bocian, M., Shiang, S. & Wasmuth, J. J. A High-resolution physical and transcript map of the Cri du Chat region of human chromosome 5p. *Genome Res.* **7**, 787–801 (1997).
5. Puechberty, J. *et al.* Genetic and physical analyses of the centromeric and pericentromeric regions of human chromosome 5: Recombination across 5cen. *Genomics* **56**, 274–287 (1999).
6. Riethman, H. C. *et al.* Integration of telomere sequences with the draft human genome sequence. *Nature* **409**, 948–951 (2001).
7. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
8. Schmutz, J. *et al.* Quality assessment of the human genome sequence. *Nature* **429**, 365–368 (2004).
9. Olivier, M. *et al.* A high-resolution radiation hybrid map of the human genome draft sequence. *Science* **291**, 1298–1302 (2001).
10. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nature Genet.* **31**, 241–247 (2002).
11. Grimwood, J. *et al.* The DNA sequence and biology of human chromosome 19. *Nature* **428**, 529–535 (2004).
12. Heilig, R. *et al.* The DNA sequence and analysis of human chromosome 14. *Nature* **421**, 601–607 (2003).
13. Hiller, L. W. *et al.* The DNA sequence of human chromosome 7. *Nature* **424**, 157–164 (2003).
14. Melki, J. *et al.* De novo and inherited deletions of the 5q13 region in spinal muscular atrophies. *Science* **264**, 1474–1477 (1994).
15. Monani, U. *et al.* A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2. *Hum. Mol. Genet.* **8**, 1177–1183 (1999).
16. Chen, Q. *et al.* Sequence of a 131-kb region of 5q13.1 containing the spinal muscular atrophy candidate genes SMN and NAIP. *Genomics* **48**, 121–127 (1998).
17. Locke, D. P. Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. *Genome Res.* **13**, 347–357 (2003).

18. Yunis, J. J. & Prakash, O. The origin of man: a chromosomal pictorial legacy. *Science* **215**, 1525–1530 (1982).
19. Noor, M. A., Grams, K. L., Bertucci, L. A. & Reiland, J. Chromosomal inversions and the reproductive isolation of species. *Proc. Natl Acad. Sci. USA* **98**, 12084–12088 (2001).
20. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
21. Groenen, M. A. *et al.* A consensus linkage map of the chicken genome. *Genome Res.* **10**, 137–147 (2000).
22. Nanda, I. *et al.* 300 million years of conserved synteny between chicken Z and human chromosome 9. *Nature Genet.* **21**, 258–259 (1999).
23. Clark, A. G. *et al.* Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* **302**, 1960–1963 (2003).
24. Hocking, L. J. *et al.* Domain-specific mutations in sequestosome 1 (SQSTM1) cause familial and sporadic Paget's disease. *Hum. Mol. Genet.* **11**, 2735–2739 (2002).
25. Pennacchio, L. A. & Rubin, E. M. Genomic strategies to identify mammalian regulatory sequences. *Nature Rev. Genet.* **2**, 100–109 (2001).
26. Ghanem, N. *et al.* Regulatory roles of conserved intergenic domains in vertebrate Dlx bigene clusters. *Genome Res.* **13**, 533–543 (2003).
27. Nobrega, M. A., Ovcharenko, I., Afzal, V. & Rubin, E. M. Scanning human gene deserts for long-range enhancers. *Science* **302**, 413 (2003).
28. Flint, J. *et al.* Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the alpha globin cluster. *Hum. Mol. Genet.* **10**, 371–382 (2001).
29. Hästbacka, J. *et al.* The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. *Cell* **78**, 1073–1087 (1994).
30. Barton, D. E., Foellmer, B. E., Wood, W. I. & Francke, U. Chromosome mapping of the growth hormone receptor gene in man and mouse. *Cytogenet. Cell Genet.* **50**, 137–141 (1989).
31. Kurotaki, N. *et al.* Haploinsufficiency of NSD1 causes Sotos syndrome. *Nature Genet.* **30**, 365–366 (2002).
32. Kurotaki, N. *et al.* Fifty microdeletions among 112 cases of Sotos syndrome: low copy repeats possibly mediate the common deletion. *Hum. Mutat.* **22**, 378–387 (2003).
33. International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
34. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using Phred. I. accuracy assessment. *Genome Res.* **8**, 175–185 (1998).
35. Gordon, D., Abajian, C. & Green, P. Consed: A graphical tool for sequence finishing. *Genome Res.* **8**, 195–202 (1998).
36. McMurray, A. A., Sulston, J. E. & Quail, M. A. Short insert libraries as a method of problem solving in genome sequencing. *Genome Res.* **8**, 562–566 (1998).
37. Schuler, G. D. Sequence mapping by electronic PCR. *Genome Res.* **7**, 541–550 (1997).
38. Jurka, J. Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet.* **16**, 418–420 (2000).
39. Zhang, Z., Harrison, P. M., Liu, Y. & Gerstein, M. Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res.* **13**, 2541–2558 (2003).
40. Bailey, J. A., Yavor, A. M., Massa, H. F., Trask, B. J. & Eichler, E. E. Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res.* **11**, 1005–1017 (2001).
41. Parsons, J. D. Miropeats: graphical DNA sequence comparisons. *Comput. Appl. Biosci.* **11**, 615–619 (1995).
42. Schwartz, S. *et al.* Human-mouse alignments with BLASTZ. *Genome Res.* **13**, 103–107 (2003).
43. Brudno, M. *et al.* LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.* **13**, 721–731 (2003).
44. Kurdoa-Kawaguchi, T. *et al.* The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nature Genet.* **29**, 279–286 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the International Chimpanzee Sequencing Consortium for pre-publication access to and permission to analyse the relevant portions of the chimpanzee genomic sequence, and the Washington University Genome Sequencing Center for pre-publication access to the chicken genomic assembly. We also thank M. Christensen, P. Butler and E. Fields for technical support, D. Gordon of the University of Washington for his assistance in developing and customizing finishing tools, T. Furey and G. Schuler for their efforts towards assessing the quality and completeness of our assembly, and P. DeJong for the construction of genomic resources. This work was performed under the auspices of the US DOE's Office of Science, Biological and Environmental Research Program, by the University of California, Lawrence Livermore National Laboratory, Lawrence Berkeley National Laboratory and Stanford University.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.S. (jeremy@shgc.stanford.edu) and E.M.R. (EMRubin@lbl.gov). The chromosome 5 sequence has been deposited in GenBank under accession number NC_000005.

Mg isotope evidence for contemporaneous formation of chondrules and refractory inclusions

Martin Bizzarro^{1,2}, Joel A. Baker^{1,3} & Henning Haack²

¹Danish Lithosphere Centre, Øster Voldgade 10, and ²Geological Museum, Øster Voldgade 5-7, DK-1350, Denmark

³School of Earth Sciences, Victoria University of Wellington, PO Box 600, Wellington, New Zealand

Primitive or undifferentiated meteorites (chondrites) date back to the origin of the Solar System¹, and thus preserve a record of the physical and chemical processes that occurred during the earliest evolution of the accretion disk surrounding the young Sun. The oldest Solar System materials present within these meteorites are millimetre- to centimetre-sized calcium-aluminium-rich inclusions (CAIs) and ferromagnesian silicate spherules (chondrules), which probably originated by thermal processing of pre-existing nebula solids²⁻⁴. Chondrules are currently believed to have formed ~2–3 million years (Myr) after CAIs (refs 5–10)—a timescale inconsistent with the dynamical lifespan of small particles in the early Solar System¹¹. Here, we report the presence of excess ²⁶Mg resulting from *in situ* decay of the short-lived ²⁶Al nuclide in CAIs and chondrules from the Allende meteorite. Six CAIs define an isochron corresponding to an initial ²⁶Al/²⁷Al ratio of $(5.25 \pm 0.10) \times 10^{-5}$, and individual model ages with uncertainties as low as $\pm 30,000$ years, suggesting that these objects possibly formed over a period as short as 50,000 years. In contrast, the chondrules record a range of initial ²⁶Al/²⁷Al ratios from (5.66 ± 0.80) to $(1.36 \pm 0.52) \times 10^{-5}$, indicating that Allende chondrule formation began contemporaneously with the formation of CAIs, and continued for at least 1.4 Myr. Chondrule formation processes recorded by Allende and other chondrites may have persisted for at least 2–3 Myr in the young Solar System.

With a half-life ($t_{1/2}$) of ~730,000 yr, the ²⁶Al to ²⁶Mg decay scheme is a high-resolution, relative chronometer for objects formed in the first 5 Myr of the Solar System^{2,5-9}. In addition, Mg isotopes can also undergo mass-dependent fractionation during evaporation and/or condensation processes, and hence provide constraints on the origin of the oldest Solar System solids. The validity of the ²⁶Al–²⁶Mg chronometer relies on the assumption that the ²⁶Al nuclide was injected, perhaps from a nearby stellar source, and homogenized within the solar nebula in a short period of time compared to its half-life¹²⁻¹⁴, and this is supported by both theoretical constraints and measurements of the coupled abundances of ²⁶Al and other extinct radionuclides in meteorites^{15,16}.

The majority of CAIs, including those from different classes of chondrites, have large ²⁶Mg excesses (²⁶Mg*) related to the decay of ²⁶Al that define an initial ²⁶Al/²⁷Al ratio ($(^{26}\text{Al}/^{27}\text{Al})_0$) of 5×10^{-5} (ref. 2), which suggests widespread distribution of the ²⁶Al nuclide in the early Solar System^{5,7,9}. In contrast, most chondrules have small ²⁶Mg excesses corresponding to $(^{26}\text{Al}/^{27}\text{Al})_0$ of $\leq 1 \times 10^{-5}$ (refs 5–9), or values that are too small to measure with most analytical methods, because Al/Mg ratios are typically much lower in chondrules than CAIs. Taken at face value, the difference in $(^{26}\text{Al}/^{27}\text{Al})_0$ recorded by the CAIs and chondrules implies that chondrules started forming ~2 Myr later than CAIs. This time gap between CAI and chondrule formation was apparently supported by high-precision Pb–Pb dating of CAIs and chondrules from chondrites¹⁰, where chondrules yielded a Pb–Pb age about 2 Myr younger than CAIs. Unfortunately, these dates are not on CAIs and chondrules from the same meteorite, so care is needed

when using these ages to understand the timing and formation processes of these oldest Solar System solids.

Further, the vast majority of existing ²⁶Al–²⁶Mg data on CAIs and chondrules are *in situ* measurements obtained by ion microprobe. Given the analytical difficulties in determining the presence of ²⁶Mg* in low Al/Mg minerals or bulk samples with *in situ* techniques, the age of the studied object is essentially defined by Al-rich phases (for example, anorthitic plagioclase) that are only a few tens to hundreds of micrometres in size. As such, chronological constraints inferred from these types of measurements could be problematical: they may be compromised and yield younger apparent ages that reflect remobilization of Mg during reprocessing of solids around the energetic young Sun or even later metamorphic events within accreted chondrite parent bodies. Using a different technique (multi-collector inductively coupled plasma mass spectrometry; MC-ICP-MS), Galy *et al.*¹⁷ recently reported high-precision Mg isotope measurements of whole chondrules from the Allende meteorite. Their study showed that at least one Al-rich chondrule formed as early as $1.4^{+0.7}_{-0.5}$ Myr after the formation of CAIs. In the light of these preliminary results, we have investigated the Mg isotope systematics of CAIs and chondrules from the Allende carbonaceous chondrite by MC-ICP-MS. Our measurements were obtained on bulk chondrules and CAIs, and are thus less susceptible to the effects of post-crystallization redistribution of ²⁶Mg* amongst different minerals than are *in situ* analyses. The external reproducibility of the ²⁶Mg* value in our laboratory inferred from repeated analyses of modern terrestrial basalts is $\pm 0.017\%$ ($n = 17$; 2 s.d.), which enables the detection of radiogenic ²⁶Mg in objects with ²⁷Al/²⁴Mg ratios ≥ 0.12 if they formed contemporaneously with CAIs.

Our new ²⁶Al–²⁶Mg data for CAIs and chondrules from the Allende carbonaceous chondrite are reported in Table 1. Five different fragments of a coarse-grained, type B CAI (A8) define a statistically significant isochron corresponding to an $(^{26}\text{Al}/^{27}\text{Al})_0$ value of $(5.34 \pm 0.51) \times 10^{-5}$; that is, indistinguishable from the canonical $(^{26}\text{Al}/^{27}\text{Al})_0$ value of 5×10^{-5} (Fig. 1a). Five additional CAIs (type A and B) have inferred $(^{26}\text{Al}/^{27}\text{Al})_0$ values indistinguishable from the canonical value (Fig. 1a) and a combined regression of all CAIs also defines a statistically significant isochron corresponding to an $(^{26}\text{Al}/^{27}\text{Al})_0$ value of $(5.25 \pm 0.10) \times 10^{-5}$. The interpretation of the correlation observed in Fig. 1a as an isochron assumes that the analysed CAIs were ultimately derived from a unique Mg and Al reservoir, and this appears to be a valid assumption given the self-consistency of the data set. These results are in agreement with previous Al–Mg investigations of CAIs (ref. 2), and validate the analytical protocols followed in this study—even in CAI material, which has low Mg compared to the analysed chondrules. Furthermore, our high-precision data indicate that these Allende CAIs formed over an extremely short time span, because the analytical uncertainty of the CAI isochron corresponds to ~40,000 yr, which is comparable to the internal precision on most of the individual CAI model ages, leaving the possibility that these objects formed within an even shorter time window. However, while model ages for all these individual CAIs are within uncertainty, the uncertainties potentially allow the oldest and youngest CAIs to have an age difference of up to 150,000 yr. *In situ* ion probe measurements of minerals with high Al/Mg ratios in CAIs from Allende, although typically characterized by large ²⁶Mg excesses, frequently yield $(^{26}\text{Al}/^{27}\text{Al})_0$ values lower than the canonical value (for example, the compilation in ref. 2). This suggests that *in situ* measurements of Mg isotopes in meteorite material may indeed be compromised by Mg remobilization, although further Mg isotope study of bulk samples of CAIs (and chondrules) for which *in situ* data are available is required to assess this hypothesis with certainty.

Fifteen chondrules with ²⁷Al/²⁴Mg ratios ranging from 0.026 to 1.08 were analysed, and resolvable ²⁶Mg* correlated with the ²⁷Al/²⁴Mg ratio was detected in 11 chondrules. The magnitude of

the excesses in the chondrules varies from 0.039 to 0.378‰ (with respect to the CAI initial $\delta^{26}\text{Mg}^*$ of -0.024‰ ; Fig. 1): that is, ~ 2 to 20 times higher than our external reproducibility (Table 1). The presence of $^{26}\text{Mg}^*$ in these chondrules might either reflect their formation ages or, alternatively, the radiogenic ^{26}Mg could have been inherited from re-melting of CAI-like precursor materials, in which case the data instead provide an indication of the degree of mixing between distinct components and have no age significance. In principle, the presence of $^{26}\text{Mg}^*$ in chondrules A-B6, A-B9 and A-C2 might be explained by incorporation of 10% to $>50\%$ of CAI material during the chondrule-forming event (Fig. 1b). The internal isochron defined by three analyses of A-B9 (inset to Fig. 1b) could thus reflect a mixing relationship, but this explanation requires that this chondrule incompletely incorporated the CAI material while it formed (from 13 to 20%), and that we have sampled different subdomains in the chondrule containing varying amounts of CAI

material. However, the stable Mg isotope ratios of the different samples of A-B9 are not correlated with the magnitude of the $^{26}\text{Mg}^*$ (Fig. 2), and this feature is not easily reconciled with incomplete mixing of varying amounts of CAI material, given the typically very light or heavy stable Mg isotopic composition of CAIs (Table 1). In fact, for most chondrules in this study, two-component mixing between CAI-like material and material akin to chondrules with low Al/Mg ratios would require the involvement of an additional component with high $\delta^{25}\text{Mg}$ values and/or high Al abundances and a $^{26}\text{Mg}^*$ corresponding to a $(^{26}\text{Al}/^{27}\text{Al})_0$ value somewhat lower than the canonical value of 5×10^{-5} , which has yet to be identified in either our data set or other high-precision Mg isotope studies¹⁷. As such, except perhaps for chondrules A-C17 and A-C2, we interpret the presence of $^{26}\text{Mg}^*$ in the remaining chondrules to be related to the *in situ* decay of ^{26}Al incorporated during the chondrule-forming event(s). Model isochrons for these chondrules yield $(^{26}\text{Al}/^{27}\text{Al})_0$

Table 1 Al-Mg isotopic data for Allende CAIs and chondrules

Type		$^{27}\text{Al}/^{24}\text{Mg}$	$\delta^{25}\text{Mg}_{\text{DSM3}}$ (‰)	$\delta^{26}\text{Mg}^*$ (‰)	$(^{26}\text{Al}/^{27}\text{Al})_0$ ($\times 10^{-5}$)	ΔT_0 (Myr)
CAIs						
A-1	B	0.249	-3.06 ± 0.14	0.070 ± 0.017	5.27 ± 0.96	$-0.01^{+0.18}_{-0.21}$
A-3c	A	1.49	-1.80 ± 0.13	0.542 ± 0.017	5.30 ± 0.19	$-0.012^{+0.037}_{-0.039}$
A-7	B	2.95	-0.78 ± 0.13	1.07 ± 0.04	5.17 ± 0.22	$0.013^{+0.043}_{-0.045}$
A-8a	B	3.80	6.79 ± 0.14	1.39 ± 0.04	5.19 ± 0.18	$0.010^{+0.036}_{-0.037}$
A-8b	B	2.76	5.64 ± 0.11	1.01 ± 0.02	5.23 ± 0.15	$0.003^{+0.030}_{-0.031}$
A-8c	B	3.82	6.39 ± 0.01	1.41 ± 0.03	5.24 ± 0.15	$0.001^{+0.030}_{-0.031}$
A-8d	B	3.28	5.74 ± 0.10	1.25 ± 0.04	5.42 ± 0.20	$-0.035^{+0.038}_{-0.040}$
A-8e	B	3.78	5.97 ± 0.07	1.42 ± 0.03	5.33 ± 0.15	$-0.018^{+0.030}_{-0.031}$
A-11	A	2.04	0.24 ± 0.02	0.715 ± 0.020	5.05 ± 0.17	$0.038^{+0.035}_{-0.036}$
A-13	A	2.03	0.34 ± 0.04	0.743 ± 0.037	5.27 ± 0.28	$-0.006^{+0.054}_{-0.056}$
Chondrules						
A-C1	Al-rich	0.866	0.77 ± 0.04	0.107 ± 0.035	2.11 ± 0.57	$0.96^{+0.25}_{-0.33}$
A-C1	Al-rich	1.08	0.72 ± 0.04	0.081 ± 0.040	1.36 ± 0.52	$1.42^{+0.34}_{-0.51}$
A-C2	Al-rich	1.03	0.67 ± 0.01	0.354 ± 0.042	5.12 ± 0.58	$0.02^{+0.11}_{-0.13}$
A-C5	BO	0.328	0.06 ± 0.03	0.015 ± 0.002	1.66 ± 0.72	$1.21^{+0.38}_{-0.60}$
A-C14	PO	0.026	-0.23 ± 0.09	-0.007 ± 0.004	—	—
A-C17	PO	0.246	-0.32 ± 0.07	0.039 ± 0.016	3.57 ± 0.97	$0.40^{+0.25}_{-0.33}$
A-C20	n.d.	0.424	0.48 ± 0.02	0.099 ± 0.003	4.05 ± 0.57	$0.27^{+0.14}_{-0.16}$
A-C23	Al-rich	0.852	0.80 ± 0.04	0.095 ± 0.030	1.95 ± 0.49	$1.04^{+0.24}_{-0.31}$
A-C24	POP	0.345	0.24 ± 0.08	0.036 ± 0.011	2.43 ± 0.69	$0.81^{+0.26}_{-0.35}$
A-C29	POP	0.082	-0.14 ± 0.01	-0.004 ± 0.011	—	—
A-C31	PO	0.307	0.24 ± 0.04	0.065 ± 0.016	4.04 ± 0.78	$0.27^{+0.19}_{-0.23}$
A-C32	BO	0.108	-0.27 ± 0.02	-0.001 ± 0.016	—	—
A-C33	POP	0.561	0.33 ± 0.03	0.097 ± 0.008	3.01 ± 0.43	$0.58^{+0.14}_{-0.16}$
A-B3	n.d.	0.117	-0.14 ± 0.06	-0.001 ± 0.026	—	—
A-B6	BO	0.274	0.27 ± 0.06	0.083 ± 0.011	5.45 ± 0.87	$-0.04^{+0.16}_{-0.18}$
A-B9	PO	0.215	0.33 ± 0.07	0.065 ± 0.020	5.77 ± 1.30	$-0.10^{+0.21}_{-0.27}$
A-B9	PO	0.241	0.36 ± 0.10	0.073 ± 0.008	5.61 ± 0.99	$-0.07^{+0.17}_{-0.20}$
A-B9	PO	0.151	0.45 ± 0.05	0.034 ± 0.003	5.36 ± 1.60	$0.02^{+0.27}_{-0.37}$

Analysed materials comprised fragments of chondrules or CAIs $\sim 0.2\text{--}1.0\text{ mg}$ in size sampled with tungsten carbide micro-drills (500–1,000 μm diameter) from polished slabs of the Allende meteorite. Samples were digested in Savillex beakers with concentrated HF-HNO_3 on a hotplate at 150°C for 48 h. Purification of Mg was achieved by cation exchange chromatography. Total procedural blanks were $<5\text{ ng}$ and negligible. The stable Mg isotopic composition ($\delta^{25}\text{Mg}$) was determined by MC-ICP-MS using sample-standard bracketing, and expressed as the per mil (‰) deviation from the isotopic composition of the DSM3 standard (ref. 29) as follows: $\delta^{25}\text{Mg} = [({}^{25}\text{Mg}/{}^{24}\text{Mg})_{\text{sample}}/({}^{25}\text{Mg}/{}^{24}\text{Mg})_{\text{DSM3}} - 1] \times 1,000$. External reproducibility of $\delta^{25}\text{Mg}$ estimated from repeated analyses of a Mg standard solution is $\pm 0.055\text{‰}$ (2 s.d.; $n = 187$). ^{26}Mg excesses ($\delta^{26}\text{Mg}^*$) are reported as the per mil deviation of internally mass-bias-corrected $^{26}\text{Mg}/^{24}\text{Mg}$ ratios (normalized to $^{25}\text{Mg}/^{24}\text{Mg} = 0.12663$ using the exponential law) of a sample from the average mass-bias-corrected $^{26}\text{Mg}/^{24}\text{Mg}$ of two bracketing standards (DSM3), with each sample being analysed in this way at least three times. This approach does not introduce any measurable artefacts in the $\delta^{26}\text{Mg}^*$, provided that the isotopic fractionation exhibited by the samples (that is, CAIs) analysed in this study was uniformly produced by a kinetic mass fractionation process. This is in agreement with experimental evidence indicating that mass-dependent isotopic fractionation resulting from partial evaporation and/or condensation processes that occur under nebular conditions is kinetic (ref. 30). External reproducibility of $\delta^{26}\text{Mg}^*$ estimated from repeated analyses of BHVO-1 and BCR-2 basalt standards is 0.017‰ (2 s.d.; $n = 17$). Some chondrule and CAI data do not replicate as precisely as the standard because insufficient material was available to perform more than three analyses or, in the case of CAIs, it was not deemed necessary to measure $\delta^{26}\text{Mg}^*$ more precisely than shown in Table 1. The integrity of the data was investigated by (1) verifying that the chemical separation procedure does not fractionate Mg isotopes, (2) ensuring that potential isobaric interference on mass ^{26}Mg (for example, CN^+ or $^{52}\text{Cr}^{2+}$) are negligible, and (3) confirming that the influence of other elements such as Al, Na, Fe and Ca on instrumental mass bias is insignificant. Al/Mg ratios were determined by MC-ICP-MS to $\pm 2\%$ using gravimetrically prepared mixed Al-Mg standard solutions. Al/Mg ratios were measured on different sample digestions of BHVO-1 and BCR-2 basaltic standards, yielding the following values, respectively: 1.64 ± 0.01 and 3.26 ± 0.05 ($n = 3$, 2 s.d.). PO, porphyritic olivine; POP, porphyritic olivine pyroxene; BO, barred olivine; Al-rich, aluminium-rich mineralogy; n.d., not determined. $(^{26}\text{Al}/^{27}\text{Al})_0$ is calculated using the initial $^{26}\text{Mg}/^{24}\text{Mg}$ defined from the isochron calculated for all CAIs, BSE and chondrules with the canonical $(^{26}\text{Al}/^{27}\text{Al})_0$ value (that is, $-0.024 \pm 0.007\text{‰}$ compared to BSE; Fig. 1). Model ages are calculated as two-point isochrons using the CAI initial $^{26}\text{Mg}/^{24}\text{Mg}$. ΔT_0 corresponds to the deviation in the $(^{26}\text{Al}/^{27}\text{Al})_0$ of an object compared to the $(^{26}\text{Al}/^{27}\text{Al})_0$ value of $(5.24 \pm 0.07) \times 10^{-5}$ defined from the isochron (Fig. 1a) expressed in Myr. Uncertainties on $(^{26}\text{Al}/^{27}\text{Al})_0$ values (x) are derived from the propagated uncertainties in $\delta^{26}\text{Mg}^*$ (y) and $^{27}\text{Al}/^{24}\text{Mg}$ (z) using: $\sigma_x = [(\sigma_y^2/(\partial x/\partial y)^2) + (\sigma_z^2/(\partial x/\partial z)^2)]^{0.5}$. Back-scattered electron images and elemental maps of selected chondrules are available as Supplementary Information.

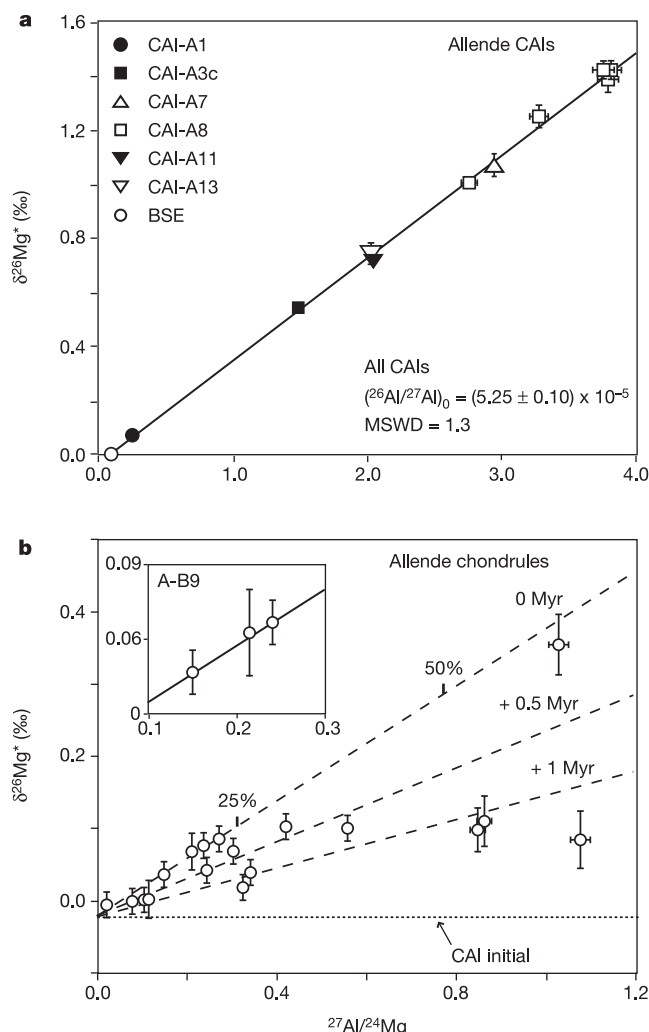


Figure 1 Al–Mg evolution diagrams. **a**, Allende CAIs. **b**, Allende chondrules. Error bars for $\delta^{26}\text{Mg}^*$ are either the 2 s.d. external reproducibility ($\pm 0.017\text{‰}$) or the 2 s.e. measured internal precision, whichever is larger (Table 1). A 2% error is assigned to the $^{27}\text{Al}/^{24}\text{Mg}$ ratio. MSWD, mean square of weighted deviations; BSE, bulk silicate earth. We interpret the intercept of the CAI regression, $-(0.024 \pm 0.016)\text{‰}$, as the initial Mg isotope composition ($^{26}\text{Mg}/^{24}\text{Mg}$) at the time of CAI formation. We note that our model BSE, defined by enstatite chondrites ($^{27}\text{Al}/^{24}\text{Mg} = 0.094 \pm 0.005$, ref. 28, and $\delta^{26}\text{Mg}^* = 0.000 \pm 0.017$), lies within error of the CAI isochron. As such, to define the initial $^{26}\text{Mg}/^{24}\text{Mg}$ precisely at the time of CAI formation, we included the BSE model and chondrules with $(^{26}\text{Al}/^{27}\text{Al})_0 = 5 \times 10^{-5}$ in the CAI regression. This yielded the following regression parameters: $(^{26}\text{Al}/^{27}\text{Al})_0 = (5.24 \pm 0.07) \times 10^{-5}$, intercept = $-(0.024 \pm 0.007)\text{‰}$, MSWD = 1.0—that is, identical to the regression parameters defined from either all CAIs, or from the CAI-8 internal isochron. The isochronous relationship of these objects with different stable Mg isotope ratios indicates that kinetic mass fractionation processes were uniformly responsible for the observed stable isotope variability. Dashed lines in **b** represent model isochrons corresponding to $T = 0$ Myr, $T = 0.5$ Myr and $T = 1$ Myr after formation of CAIs. The $T = 0$ Myr model isochron illustrates mixing between a chondrule with the lowest Al/Mg ratio and CAI-8 (off-scale), with tickmarks indicating percentage of CAI material incorporated into a hypothetical mixture between the two components. The inset to **b** shows the internal isochron for chondrule A-B9, corresponding to an $(^{26}\text{Al}/^{27}\text{Al})_0 = 5.66 \pm 0.80 \times 10^{-5}$ (using the CAI initial $^{26}\text{Mg}/^{24}\text{Mg}$ in the regression, MSWD = 0.1).

values ranging from $(5.66 \pm 0.80) \times 10^{-5}$ to $(1.36 \pm 0.52) \times 10^{-5}$ (Fig. 1b), which are the highest $(^{26}\text{Al}/^{27}\text{Al})_0$ values reported for either ferromagnesian (see Supplementary Information) or aluminium-rich chondrules. Such high abundances of ^{26}Al in chondrules support the hypothesis that ^{26}Al decay may have been the

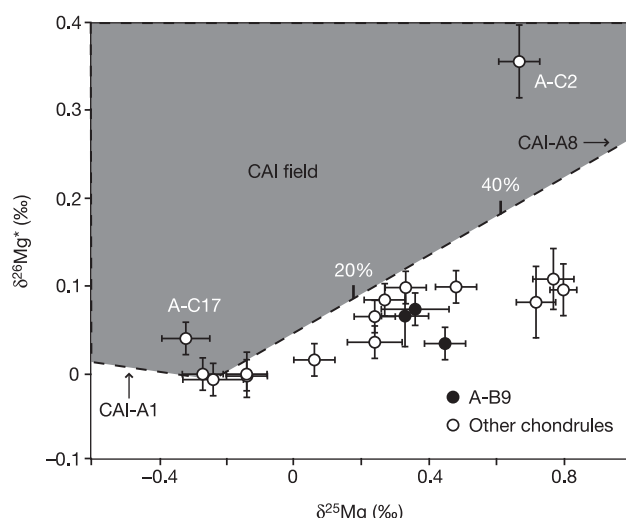


Figure 2 Mg stable isotope composition ($\delta^{26}\text{Mg}$) versus excess ^{26}Mg ($\delta^{26}\text{Mg}^*$) diagram. The error bars are either the 2 s.d. external reproducibility or the 2 s.e. measured internal precision, whichever is larger (Table 1). The dashed lines are mixing arrays between CAIs from this study with the lightest (CAI-A1) and heaviest (CAI-A8; average of five measurements) $\delta^{26}\text{Mg}$, and the chondrule with the lowest measured $^{27}\text{Al}/^{24}\text{Mg}$ ratio (Table 1). Tickmarks indicating percentage of CAI in the mixture between CAI-A8 and the chondrule with the lowest measured $^{27}\text{Al}/^{24}\text{Mg}$ ratio are shown. Two chondrules (A-C17 and A-C2) plot within the CAI-field (grey area) and may have inherited excess ^{26}Mg from precursor CAI-like material. In this case, the Al–Mg systematics of these two objects would provide no chronological information. However, as explained in the text, two-component mixing between CAI-like material and chondrules with low Al/Mg ratios cannot readily explain the presence of radiogenic ^{26}Mg in the majority of the chondrules.

major heat source for melting and differentiation in the young Solar System^{18,19}.

Assuming a homogeneous $(^{26}\text{Al}/^{27}\text{Al})_0$ in the early solar nebula, our high-precision Mg isotope data indicate that Allende chondrules started forming at the same time as the oldest CAIs, and continued for at least $1.42^{+0.51}_{-0.34}$ Myr after CAI formation. These results are consistent with occurrences of compound CAI–chondrule objects in chondrites that indicate contemporaneous formation of CAIs and chondrules^{20,21}. Recent Pb–Pb dating of chondrules from Allende²² appears to confirm our results from ^{26}Al – ^{26}Mg dating, with Allende chondrules having an absolute Pb–Pb age within error of CAIs, although the age resolution from this absolute dating is almost an order of magnitude larger than the results presented here. These results, however, are at odds with previous Pb–Pb dating of chondrules and CAIs from different meteorites¹⁰.

Combined with recently published Al–Mg data of chondrules from different types of chondrites^{6–9,17}, our study indicates that chondrule-forming event(s) may have lasted as long as 2.5 Myr following formation of CAIs. Estimates of the solar nebula time-scale, based on observational or isotopic constraints, suggest that nebula dissipation must have occurred within ~ 3 Myr of Solar System formation^{23,24}. This is consistent with both chondrules and CAIs being nebula products and, as such, our new constraints for the timing of chondrule-forming event(s) are compatible with either the x-wind model⁴ or shock wave models^{25,26} of chondrule formation. Alternatively, a protracted history for chondrule formation could also be reconciled with an origin as ejecta from colliding molten planetesimals in the accretion disk²⁷, providing that accumulation and melting of asteroids occurred < 0.2 Myr after CAI formation (considering the uncertainty of the oldest chondrule reported in the present study). Recurrent formation of chondrules over nearly 3 Myr might indicate that a number of distinct processes

and/or heat sources were involved in the formation history of these objects. □

Received 21 May; accepted 21 July 2004; doi:10.1038/nature02882.

- Alexander, C. M. O'D., Boss, A. P. & Carlson, R. W. The early evolution of the inner solar system: A meteoritic perspective. *Science* **293**, 64–68 (2001).
- MacPherson, G. J. in *Meteorites, Comets and Planets* (ed. Davis, A. M.) 201–246, Vol. 1 of *Treatise on Geochemistry* (eds Holland, H. D. & Turekian, K. K.) (Elsevier-Pergamon, Oxford, 2003).
- Rubin, A. E. Petrologic, geochemical and experimental constraints on models of chondrule formation. *Earth Sci. Rev.* **50**, 3–27 (2000).
- Shu, F. H., Shang, H., Gounelle, M., Glassgold, A. E. & Lee, T. The origin of chondrules and refractory inclusions in chondritic meteorites. *Astrophys. J.* **548**, 1029–1050 (2001).
- Russell, S. S., Srinivasan, G., Huss, G. R., Wasserburg, G. J. & MacPherson, G. J. Evidence for widespread ^{26}Al in the solar nebula and constraints for nebula time scales. *Science* **273**, 757–762 (1996).
- Kita, N. T., Nagahara, S. & Morishita, Y. A short duration of chondrule formation in the solar nebula: Evidence from Al-26 in Semarkona ferromagnesian chondrules. *Geochim. Cosmochim. Acta* **64**, 3913–3922 (2000).
- Huss, G. R., MacPherson, G. J., Wasserburg, G. J., Russell, S. S. & Srinivasan, G. Aluminum-26 in calcium-aluminum-rich inclusions and chondrules from unequilibrated ordinary chondrites. *Meteorit. Planet. Sci.* **36**, 975–997 (2001).
- Mostefaoui, S. et al. The relative formation ages of ferromagnesian chondrules inferred from their initial aluminum-26/aluminum-27 ratios. *Meteorit. Planet. Sci.* **37**, 421–438 (2002).
- Hsu, W. B., Huss, G. R. & Wasserburg, G. J. Al-Mg systematics of CAIs, POI, and ferromagnesian chondrules from Ningqiang. *Meteorit. Planet. Sci.* **38**, 35–48 (2003).
- Amelin, Y., Krot, A. N., Hutcheon, I. D. & Ulyanov, A. A. Lead isotopic ages of chondrules and calcium-aluminum-rich inclusions. *Science* **297**, 1678–1683 (2002).
- Adachi, I., Hayashi, C. & Nakazawa, K. Gas drag effect on elliptic motion of a solid body in primordial solar nebula. *Prog. Theor. Phys.* **56**, 1756–1771 (1976).
- Busso, M., Gallino, R. & Wasserburg, G. J. Nucleosynthesis in asymptotic giant branch stars: Relevance for Galactic enrichment and solar system formation. *Annu. Rev. Astron. Astrophys.* **37**, 239–309 (1999).
- Goswami, J. N. Short-lived nuclides in the early solar system: the stellar connection. *New Astron. Rev.* **48**, 125–132 (2004).
- Sahijpal, S., Goswami, J. N., Davis, A. M., Grossman, L. & Lewis, R. S. A stellar origin for the short-lived nuclides in the early Solar System. *Nature* **391**, 559–561 (1998).
- Wasserburg, G. J., Gallino, R., Busso, M., Goswami, J. N. & Raiteri, C. M. Injection of freshly synthesized ^{41}Ca in the early solar nebula by an asymptotic giant branch star. *Astrophys. J.* **440**, L101–L104 (1995).
- Marhas, K. K., Goswami, J. N. & Davis, A. M. Short-lived nuclides in hibonite grains from Murchison: Evidence for Solar System evolution. *Science* **298**, 2182–2185 (2002).
- Galy, A., Young, E. D., Ash, R. D. & O'Nions, R. K. The formation of chondrules at high gas pressures in the solar nebula. *Science* **290**, 1751–1753 (2000).
- Urey, H. C. The cosmic abundances of potassium, uranium, and thorium and the heat balances of the Earth, the Moon and Mars. *Proc. Natl Acad. Sci. USA* **41**, 127–144 (1955).
- Srinivasan, G., Goswami, J. N. & Bhandari, N. ^{26}Al in eucrite Piplia Kalan: plausible heat source and formation chronology. *Science* **284**, 1348–1350 (1999).
- Bischoff, A. & Keil, K. Al-rich objects in ordinary chondrites: Related origin of carbonaceous and ordinary chondrites and their constituents. *Geochim. Cosmochim. Acta* **48**, 693–709 (1984).
- Itoh, S. & Yurimoto, H. Contemporaneous formation of chondrules and refractory inclusions in the early Solar System. *Nature* **423**, 728–731 (2003).
- Amelin, Y., Krot, A. & Twelker, E. Pb isotopic age of the CB chondrite Gujba, and the duration of the chondrule formation interval. *Geochim. Cosmochim. Acta* **68** Abstr. E958 (2004).
- Briceño, C. et al. The CIDA-QUEST large-scale survey of Orion OB1: Evidence for rapid disk dissipation in a dispersed stellar population. *Science* **291**, 93–96 (2001).
- Podosek, F. A. & Cassen, P. Theoretical, observational, and isotopic estimates of lifetime of the solar nebula. *Meteoritics* **29**, 6–25 (1994).
- Weidenschilling, S. J., Marzari, F. & Hood, L. L. The origin of chondrules at jovian resonances. *Science* **279**, 681–684 (1998).
- Desch, S. J. & Connolly, H. C. A model of the thermal processing of particles in solar nebula shocks: Application to the cooling rates of chondrules. *Meteorit. Planet. Sci.* **37**, 183–207 (2002).
- Sanders, I. A. in *Chondrules and the protoplanetary disk* (eds Hewins, R. H., Jones, R. H. & Scott, E. R. D.) 327–334 (Cambridge Univ. Press, Cambridge, 1996).
- McDonough, W. F. & Sun, S.-S. Composition of the Earth. *Chem. Geol.* **120**, 223–253 (1995).
- Galy, A. et al. Magnesium isotope heterogeneity of the isotopic standard SRM980 and new reference materials for magnesium-isotope-ratio measurements. *J. Anal. Atomic Spectrometry* **18**, 1352–1356 (2003).
- Richter, F. M., Davis, A. M., Ebel, D. S. & Hashimoto, A. Elemental and isotopic fractionation of type B calcium-, aluminum-rich inclusions: Experiments, theoretical considerations, and constraints on their thermal evolution. *Geochim. Cosmochim. Acta* **66**, 521–540 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements Financial support for this project was provided by the Danish Lithosphere Centre (funded by the Danish National Science Foundation). M.B. is grateful for financial support in the form of an NSERC postdoctoral fellowship. We thank A. Halliday for his review of this paper.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.B. (mbi@dlc.ku.dk).

Exceptional astronomical seeing conditions above Dome C in Antarctica

Jon S. Lawrence¹, Michael C. B. Ashley¹, Andrei Tokovinin² & Tony Travoignon¹

¹School of Physics, University of New South Wales, New South Wales 2052, Australia

²Cerro-Tololo Inter American Observatory, Casilla 603, La Serena, Chile

One of the most important considerations when planning the next generation of ground-based optical astronomical telescopes is to choose a site that has excellent 'seeing'—the jitter in the apparent position of a star that is caused by light bending as it passes through regions of differing refractive index in the Earth's atmosphere. The best mid-latitude sites have a median seeing ranging from 0.5 to 1.0 arcsec (refs 1–5). Sites on the Antarctic plateau have unique atmospheric properties that make them worth investigating as potential observatory locations. Previous testing at the US Amundsen-Scott South Pole Station has, however, demonstrated poor seeing, averaging 1.8 arcsec (refs 6, 7). Here we report

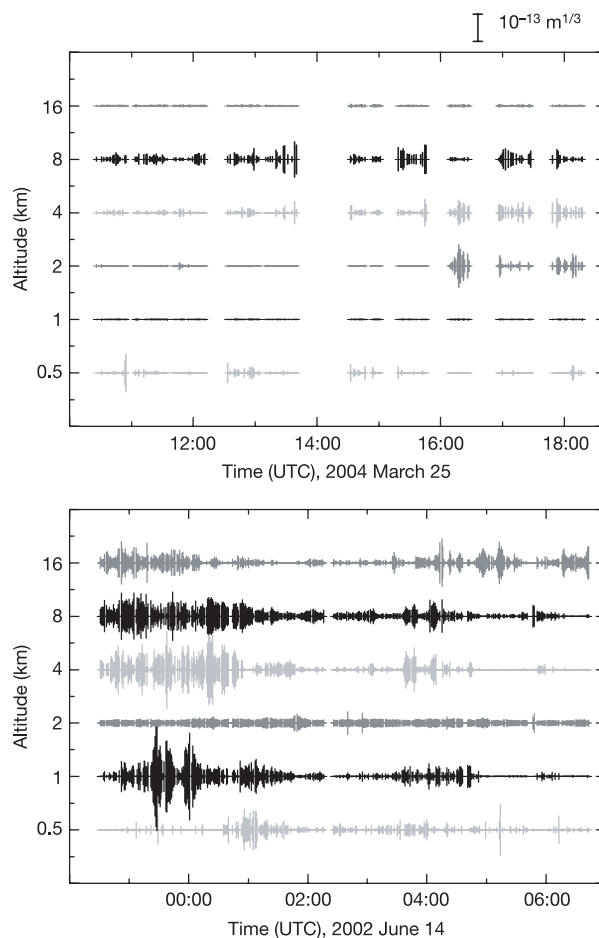


Figure 1 Comparison of turbulence profiles obtained from similar instruments at Dome C and Cerro Tololo. A typical night (representing ~50th percentile conditions) of the refractive index structure constant profile for Cerro Tololo²⁸ (bottom panel) and Dome C (top panel). Data are from similar MASS instruments. The length of the vertical bars at each altitude represents the magnitude of the refractive index structure constant integrated over that layer.

observations of the wintertime seeing from Dome C (ref. 8), a high point on the Antarctic plateau at a latitude of 75° S. The results are remarkable: the median seeing is 0.27 arcsec, and below 0.15 arcsec 25 per cent of the time. A telescope placed at Dome C would compete with one that is 2 to 3 times larger at the best mid-latitude observatories, and an interferometer based at this site could work on projects that would otherwise require a space mission.

In searching for sites for a major observatory, many factors need to be considered, including the atmospheric turbulence, cloud cover, precipitable water vapour, thermal emission from the atmosphere, auroral activity, aerosol/dust pollution, average and maximum wind speeds, seismic activity, rates of snow/rain fall, light pollution, accessibility, infrastructure and cost of operation. It has long been recognized that sites on the Antarctic plateau excel in many of these characteristics^{9–11}. However, the poor seeing at the South Pole itself, caused by a highly turbulent layer of air within 200–300 m of the ground¹², is a major limitation. It has been postulated that such a turbulent layer may be absent at Dome C, owing to the local topography, lower wind speeds and higher altitude (3,250 m, compared to 2,840 m)^{13,14}. Seeing measurements from Dome C during the 2003–04 summertime (daytime in Antarctica) have eliminated sources of local turbulence that had affected earlier measurements¹⁵ and have now demonstrated periods of superb seeing below 0.2 arcsec (E. Aristidi and E. Fossat, personal communication). Summertime balloon-borne experiments have shown low levels of high-altitude winds¹⁶. However, the crucial information that astronomers need, and that we provide here, are measurements of the seeing in the wintertime, after sunset. Obtaining this data presented a formidable technical challenge, given that the French/Italian station at Dome C is currently uninhabited during winter⁸, and that infrastructure such as electrical power and communications is not available.

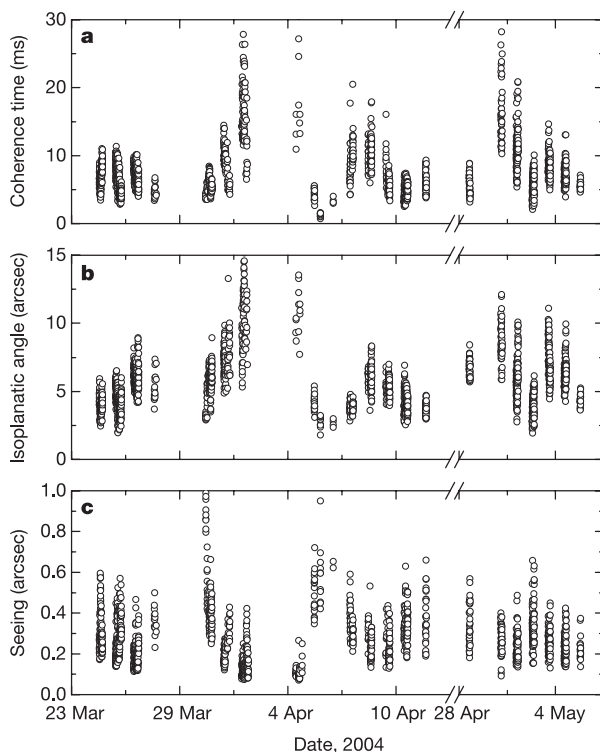


Figure 2 Dome C atmospheric coherence time, isoplanatic angle and seeing data as a function of time. Coherence time (a) and isoplanatic angle (b) are derived from the MASS instrument. Atmospheric seeing above 30 m (c) is computed from a combination of the refractive index structure constants from MASS and SODAR. Data covers the period 23 March to 5 May 2004. All data refer to observations at 1–1.5 airmass scaled to the zenith, and at a wavelength of 500 nm.

To obtain data over the winter months, we developed a remote autonomous laboratory, the AASTINO (Automated Astrophysical Site Testing International Observatory)¹⁷. This observatory was constructed at Dome C in January 2003, and it provides heat, electrical power, Iridium satellite communications, and computer control for a series of site testing instruments. In January 2004, we installed a Multi-Aperture Scintillation Sensor (MASS)¹⁸ to measure the wintertime seeing. MASS uses the spatial/temporal structure of single star scintillation (that is, intensity fluctuations) to evaluate vertical refractive index fluctuation profiles^{19,20}. The advantage of MASS over other techniques for measuring the seeing is that MASS uses a small telescope (making it less costly, and easier to automate) and it is able to measure the contributions to the seeing from six layers within the atmosphere, at fixed altitudes of 0.5, 1, 2, 4, 8 and 16 km. A limitation of MASS is that it is insensitive to seeing below about 500 m. We have therefore simultaneously used a SODAR (Sonic Detection And Ranging) instrument to determine the contribution from the layer between 30 and 500 m. From the MASS profiles, the ‘free-atmosphere’ seeing (above 500 m) and the isoplanatic angle (the angle over which atmospheric phase fluctuations are coherent) can be determined, and have been well-verified against other techniques (J. Vernin, A. Ziad and A.T., manuscript in preparation). The atmospheric coherence time (defined as the time over which the phase fluctuations are coherent) can also be derived from MASS, although the validity of this derivation for a complete range of atmospheric conditions is yet to be fully confirmed²¹. A MASS is currently in operation at Cerro Tololo¹⁹ and several more are being deployed to other sites as part of a global site-testing effort.

The Dome C SODAR has been calibrated against microthermal sensors²². Robust performance of this instrument under Antarctic winter conditions, and good agreement with DIMM (Differential Image Motion Monitor) measurements, have previously been demonstrated through operation at the South Pole¹². The calibration of the SODAR instrument is also significantly simplified in the low absolute humidity environment on the Antarctic plateau.

Figure 1 compares a typical night’s profile of the refractive index structure constant (which describes the refractive index variations) from the Dome C MASS and a similar MASS instrument at the Cerro Tololo Inter American Observatory in Chile. The Tololo atmosphere, typical of mid-latitude sites, exhibits strong turbulence within the lower troposphere, extending up to 1 km above the surface. Additionally, strong turbulence is observed throughout the upper troposphere bounded by the jet stream at 10–14 km above the surface. The Dome C turbulence profile is quite different, and is unlike any observed at mid-latitude sites. The strongest turbulent

Table 1 Comparison of observatory site conditions

Site	ϵ_0	θ_0	τ_0
Dome C	0.27	5.7	7.9
South Pole	1.8	3.2	1.6
Mauna Kea	0.5–0.7	1.9	2.7
San Pedro Martir	0.59	1.6	6.5
Cerro Paranal	0.80	2.6	3.3
La Palma	0.76	1.3	6.6

ϵ_0 is the seeing in arcseconds, θ_0 is the isoplanatic angle in arcseconds and τ_0 is the coherence time in milliseconds. All values are corrected to the zenith and at a wavelength of 500 nm. Seeing, isoplanatic angle and coherence time at South Pole are mean total atmosphere values (above ground-level) from 16 microthermal balloon launches in winter 1995 combined with microthermal tower measurements of the 0–30 m ground layer⁷. Seeing and isoplanatic angle values at Mauna Kea, Hawaii are based on 20 nights of SCIDAR observations in 1995 (seeing above ground level)¹, and FWHM measurements from the Auto Guider of the Subaru telescope during focus checks over a 12 month period from 2000 to 2001 (seeing above ~15 m)⁴. Seeing from San Pedro Martir, Mexico, is the median from 2 yr of DIMM measurements (seeing above 8 m)⁵. Isoplanatic angle and coherence time from San Pedro Martir are obtained with generalized SCIDAR (located 15 m above ground level) over 27 nights in 1997 and 2000²⁹. Seeing and isoplanatic angle from Cerro Paranal, Chile, are average values from DIMM measurements (above 5 m) over 10 yr (1989–95 and 1998–2002)². The coherence time at Cerro Paranal is derived (to an accuracy of 20%) from DIMM measurements combined with balloon-borne wind speed measurements². Seeing from La Palma, Canary Islands, is from 9 months of DIMM measurements (seeing above 5 m)³. Isoplanatic angle and coherence time from La Palma are from six microthermal balloon launches³⁰ in 1990.

layer occurs at a lower altitude owing to the lower tropopause height (5–8 km above the surface), and is of lower intensity owing to the lack of strong winds at this altitude. The only mechanism for turbulence generation in the Antarctic stratosphere above 10 km is from the Antarctic polar vortex, a system of strong high-altitude winds circling the continent²³. Prior to our MASS data, it was a matter for speculation to what extent the vortex winds would affect the turbulence at Dome C. Our data show no evidence thus far of such turbulence.

The low values of high-altitude turbulence at Dome C lead to refractive index variations that are coherent over a large angle (that is, large isoplanatic angle) and over a long time (that is, large coherence time), see Fig. 2. The average values, 5.7 arcsec and 7.9 ms respectively, are higher than average values recorded at any other site, as shown in Table 1.

Combining the refractive index structure constant values for the free atmosphere determined by the MASS, with the surface boundary layer turbulence determined by the SODAR, gives the atmospheric seeing above 30 m (Fig. 2, Table 1). Whereas the surface boundary layer (within a few hundred metres of ground level) is usually a significant contribution to the total atmospheric seeing at mid-latitude sites, the 30–900 m turbulence observed at Dome C by the SODAR is exceptionally low, and is typically below the SODAR detection threshold (0.05 arcsec). This is expected from the very low-velocity ground level winds at this site. An additional contribution to the total atmospheric seeing is expected from the first 30 m above ground level (a region in which the SODAR is not sensitive).

A histogram of the atmospheric seeing above 30 m is shown in Fig. 3, and is compared to the probability distribution for seeing at Mauna Kea and Cerro Paranal. The mean Dome C seeing of 0.27 arcsec is only observed at mid-latitude sites under exceptionally calm conditions ($\ll 1\%$ of the time). The best seeing we observed at Dome C was 0.07 arcsec, which, to our knowledge, represents the lowest value reported anywhere.

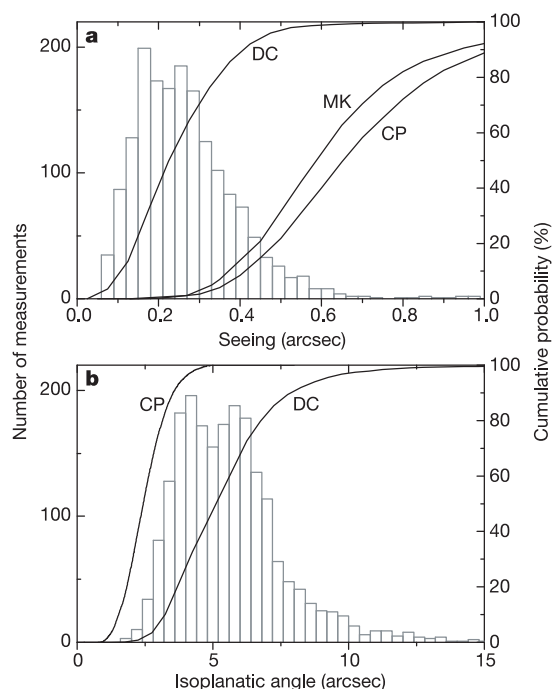


Figure 3 Histograms and cumulative distributions of the atmospheric seeing and the isoplanatic angle. **a**, Histogram of Dome C seeing above 30 m from MASS combined with SODAR, and cumulative distributions of seeing at Dome C (DC), Mauna Kea (MK) (derived from ref. 4), and Cerro Paranal (CP)². **b**, Histogram of Dome C isoplanatic angle derived from the MASS instrument, and the cumulative distribution of isoplanatic angle from Dome C and Cerro Paranal².

The atmospheric characteristics of a site strongly influence the degree of correction and field of view achievable by using adaptive optics (AO)^{24,25}. A factor of 4–10 fewer actuators would be required on a Dome C AO system to achieve the same residual wavefront error as a mid-latitude system. The larger isoplanatic angle of the Dome C atmosphere leads to a factor of ~ 3 increase in the field of view correctable by AO. The longer atmospheric time constant allows increased integration times for the AO wavefront sensor, which means that fainter guide stars can be used. The combination of a long coherence time with a large isoplanatic angle results in a greatly increased sky coverage for an AO system, both with natural and laser guide stars²⁴. An AO system at Dome C would thus provide a higher level of correction for a larger fraction of the sky, compared to any other site. Additionally, for multi-conjugate AO, the complexity (number of deformable mirrors and actuators per mirror) is significantly reduced.

A world-wide search is currently being conducted to determine the most appropriate sites for the next generation of large optical and infrared telescopes. Advantages of the Dome C atmosphere for astronomy include the very low infrared sky emission, 10–100 times lower than observed from any mid-latitude site; the high percentage ($>75\%$) of cloud-free time²⁶, which is comparable with the best mid-latitude sites; the low atmospheric precipitable water vapour content, which in winter should be lower than at any other site so far investigated, resulting in significant increases in atmospheric transmission²⁷; and the low aerosol and dust content of the atmosphere. Advantageous site conditions include the very low ground level wind speeds¹⁶ and lack of seismic activity (which reduces structural requirements on telescope mounts and domes), and low levels of light pollution. These advantages of the Dome C site must be weighed against its accessibility, any associated increase in system cost resulting from this, and any engineering issues resulting from the extremely low site temperatures. Many of the benefits of Dome C are also characteristics of the other Antarctic plateau stations. The poor ground-level seeing found at the South Pole station, however, severely limits its applicability for optical astronomy. Although it is expected that the turbulence conditions at Dome A, the highest point on the Antarctic plateau at an altitude of 4,200 m, will be superior even to Dome C, the complete lack of infrastructure at this site (it has never been visited) means that Dome C may be a preferable location.

The extremely favourable seeing, the large isoplanatic angle, and the long atmospheric coherence time reported here are compelling advantages that lead us to conclude that Dome C is the best ground-based site to develop a new astronomical observatory. $\square$

Received 19 May; accepted 9 August 2004; doi:10.1038/nature02929.

1. Racine, R. & Ellerbroek, B. L. Profiles of the night-time turbulence above Mauna Kea and isoplanatism extension in adaptive optics. *Proc. SPIE* **2534**, 248–257 (1995).
2. Sarazin, M. Astroclimatology of Paranal (European Southern Observatory, 1999); (<http://www.eso.org/gen-fac/pubs/atclim/paranal/index.html>) (1999).
3. Michel, R., Echevarria, J., Costero, R. & Harris, O. The seeing at San Pedro Mártir Observatory as measured using the DIMM method. *Rev. Mex. Astron. Astrofis.* **19**, 37–40 (2003).
4. National Astronomical Observatory of Japan. Subaru telescope seeing (<http://www.naoj.org/Observing/Telescope/Image/seeing.html>) (2003).
5. Muñoz-Tuñón, C., Vernin, J. & Varela, A. M. Night-time image quality at Roque de los Muchachos Observatory. *Astron. Astrophys. Suppl. Ser.* **125**, 183–193 (1997).
6. Travouillon, T. et al. Automated Shack-Hartmann seeing measurements at the South Pole. *Astron. Astrophys.* **409**, 1169–1173 (2003).
7. Marks, R. D., Vernin, J., Azouit, M., Manigault, J. F. & Clevelin, C. Measurement of optical seeing on the high Antarctic plateau. *Astron. Astrophys. Suppl. Ser.* **134**, 161–172 (1999).
8. Candidi, M. & Lori, A. Status of the Antarctic base at Dome C and perspectives for Astrophysics. *Mem. Soc. Astron. Ital.* **74**, 29–37 (2003).
9. Ashley, M. C. B. et al. South Pole observations of the near-infrared sky brightness. *Pub. Astron. Soc. Pacif.* **108**, 721–723 (1996).
10. Chamberlain, M. A. et al. Mid-infrared observing conditions at the South Pole. *Astrophys. J.* **535**, 501–511 (2000).
11. Chamberlain, R. A. & Bally, J. 225-GHz atmospheric opacity of the South Pole sky derived from continual radiometric measurements of the sky-brightness temperature. *Appl. Opt.* **33**, 1095–1099 (1994).
12. Travouillon, T., Ashley, M. C. B., Burton, M. G., Storey, J. W. V. & Lowenstein, R. F. Atmospheric turbulence at the South Pole and its implications for astronomy. *Astron. Astrophys.* **400**, 1163–1172 (2002).

13. Gillingham, P. R. Super seeing from the Australian Antarctic Territory? *ANARE Res. Notes* **88**, 290–292 (1993).
14. Marks, R. D. Astronomical seeing from the summits of the Antarctic plateau. *Astron. Astrophys.* **385**, 328–336 (2002).
15. Aristidi, E. *et al.* Antarctic site testing: first daily seeing monitoring at Dome C. *Astron. Astrophys.* **406**, L19–L22 (2003).
16. Aristidi, E. *et al.* Preliminary summer site testing study based on weather balloons at Dome C, Antarctica. *Astron. Astrophys.* (submitted).
17. Lawrence, J. S., Ashley, M. C. B. & Storey, J. W. V. A remote, autonomous laboratory for Antarctica with hybrid power generation. *J. Electron. Electric. Eng. Aust.* (in the press).
18. Kornilov, V. *et al.* MASS: a monitor of the vertical turbulence distribution. *Proc. SPIE* **4839**, 837–845 (2003).
19. Tokovinin, A., Baumont, S. & Vasquez, J. Statistics of turbulence profile at Cerro Tololo. *Mon. Not. R. Astron. Soc.* **340**, 52–58 (2003).
20. Tokovinin, A., Kornilov, V., Shatsky, N. & Vozniakova, O. Restoration of turbulence profile from scintillation indices. *Mon. Not. R. Astron. Soc.* **343**, 891–899 (2003).
21. Tokovinin, A. Measurement of seeing and atmospheric time constant by differential scintillations. *Appl. Opt.* **41**, 957–964 (2002).
22. Travouillon, T., Burton, M. G., Storey, J. W. V., Vernin, J. & Azouit, M. Comparison and cross-calibration of SODAR and microthermal sensor results. *Geophys. Res. Lett.* (submitted).
23. Chanin, M. An exceptional situation in the Antarctic stratosphere in 2002. *Mem. Soc. Astron. Ital. Suppl.* **2**, 19–25 (2003).
24. Ellerbroek, B. L. & Tyler, D. W. Adaptive optics sky coverage calculations for the Gemini-North telescope. *Publ. Astron. Soc. Pacif.* **110**, 165–185 (1998).
25. Beckers, J. M. Adaptive optics for astronomy: principles, performance, and applications. *Annu. Rev. Astron. Astrophys.* **31**, 13–62 (1993).
26. Ashley, M. C. B., Burton, M. G., Calisse, P. G., Phillips, A. & Storey, J. W. V. Site testing at Dome C—cloud statistics from the ICECAM experiment. *Highlights of Astronomy* (ASP Conf. Ser. Vol. 13, Astronomical Society of the Pacific, in the press).
27. Lawrence, J. S. Infrared and submillimeter atmospheric characteristics of high Antarctic Plateau sites. *Publ. Astron. Soc. Pacif.* **116**, 482–492 (2004).
28. National Optical Astronomy Observatory. MASS data access (<http://139.229.11.21/massindex.php?Submit2=MASS+Database>) (2004).
29. Avila, R. *et al.* Optical turbulence and wind profiles at San Pedro Martir. *Rev. Mex. Astron. Astrofis.* **19**, 11–22 (2003).
30. Vernin, J. & Muñoz-Tuñón, C. Optical seeing at La Palma Observatory. II. Intensive site testing campaign at the Nordic Optical Telescope. *Astron. Astrophys.* **284**, 311–318 (1994).

Acknowledgements The AASTINO project is supported by the Australian Research Council, the Australian Antarctic Division, the French and Italian Antarctic research programs (IPEV, PNRA) and the US National Science Foundation. The Dome C MASS is a collaboration between UNSW, CTIO and JPL. We thank J. Storey, A. Moore and C. Bonner for contributions during Dome C deployments; J. Storey, M. Swain, J. Lloyd, and J. Everett for intellectual input into the Dome C MASS project; S. Kenyon for assistance with data analysis; and V. Kornilov and N. Shatsky for software support of MASS. We also thank P. Gillingham for his encouragement over the past decade to demonstrate the existence of exceptional seeing from the Antarctic plateau.

Authors' contributions J.S.L. and M.C.B.A. designed and operated the Antarctic MASS instrumentation and co-wrote this Letter. A.T. provided the MASS sensor in collaboration with V. Kornilov *et al.* and assisted with the data analysis. T.T. assisted with the SODAR data and installation of MASS at Dome C.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.C.B.A. (m.ashley@unsw.edu.au).

Magnetic trapping of rare-earth atoms at millikelvin temperatures

Cindy I. Hancox, S. Charles Doret, Matthew T. Hummon, Linjiao Luo & John M. Doyle

Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA

The ability to create quantum degenerate gases has led to the realization of Bose–Einstein condensation of molecules^{1–4}, atom–atom entanglement⁵ and the accurate measurement of the Casimir force in atom–surface interactions⁶. With a few exceptions^{7–9}, the achievement of quantum degeneracy relies on evaporative cooling of magnetically trapped atoms to ultracold temperatures. Magnetic traps confine atoms whose electronic

magnetic moments are aligned anti-parallel to the magnetic field. This alignment must be preserved during the collisional thermalization of the atomic cloud. Quantum degeneracy has been reached in spherically symmetric, S-state atoms (atoms with zero internal orbital angular momentum). However, collisional relaxation of the atomic magnetic moments of non-S-state atoms (non-spherical atoms with non-zero internal orbital angular momentum) is thought to proceed rapidly. Here we demonstrate magnetic trapping of non-S-state rare-earth atoms, observing a suppression of the interaction anisotropy in collisions. The atoms behave effectively like S-state atoms because their unpaired electrons are shielded by two outer filled electronic shells that are spherically symmetric. Our results are promising for the creation of quantum degenerate gases with non-S-state atoms, and may facilitate the search for time variation of fundamental constants^{10–12} and the development of a quantum computer with highly magnetic atoms¹³.

Capturing atoms in a magnetic trap requires precooling to millikelvin temperatures. Laser cooling, the dominant method for precooling, has been used for atoms with strong cycling transitions at technologically accessible wavelengths. An alternative method is buffer-gas cooling, which relies upon collisional thermalization with a cold helium buffer gas. Although collisions with buffer gas atoms drive thermalization and subsequent trapping of the atomic cloud, they may also induce reorientation of the magnetic moment of trapped atoms (Zeeman relaxation). As only atoms in the higher-energy low-field-seeking states are trapped, the Zeeman relaxation leads to trap loss.

Evaporative cooling and buffer-gas loading have so far been limited to S-state atoms. The electronic density distribution of S-state atoms is spherically symmetric and the electrostatic interaction between atoms is not effective in driving relaxation of the highest-energy Zeeman levels in S-state atom collisions. However, if one or both of the colliding atoms have non-zero electronic orbital angular momentum (non-S-state atoms), the electronic interaction between the atoms is anisotropic¹⁴. The atomic angular momentum can be strongly coupled to the rotational motion of the collision complex and it has been shown that for the non-S-state atoms

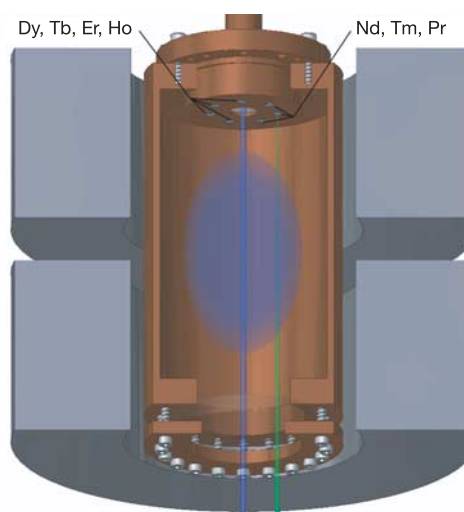


Figure 1 The experimental cell. The copper cell is thermally anchored to a dilution refrigerator. Two superconducting solenoids surround the cell. Their currents may be run in same direction to produce a uniform magnetic field, or in opposite directions to produce a spherical quadrupole trap with a depth of up to 3.5 T. The cell is filled with a buffer gas of He atoms at a density of $8 \times 10^{15} \text{ cm}^{-3}$. The rare-earth atoms are produced by laser ablation of elemental foil targets mounted at the top of the cell. The atoms are detected via absorption spectroscopy.

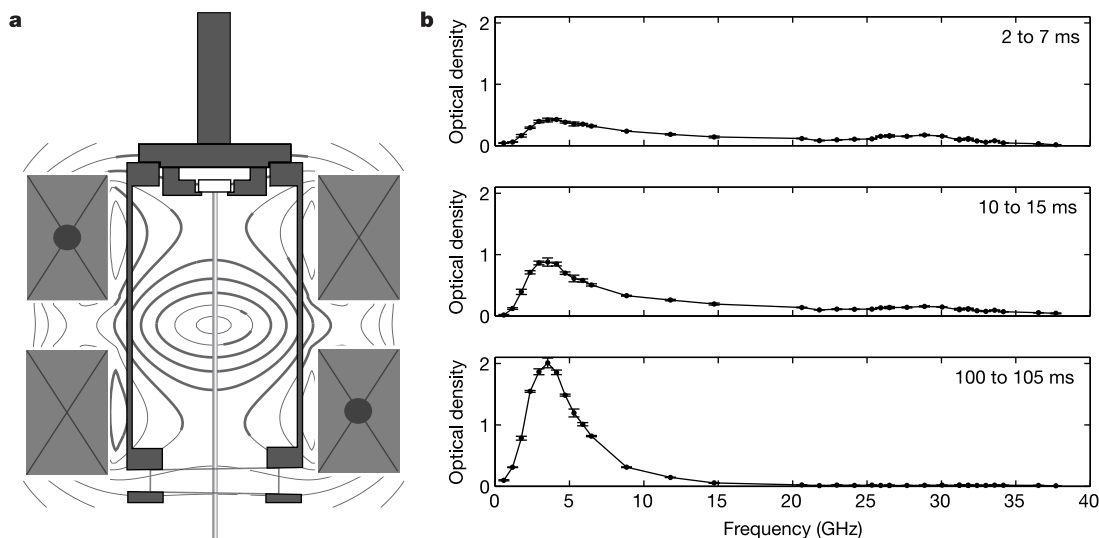


Figure 2 Time progression of trapping. **a**, The anti-Helmholtz field profile. The current in the two solenoids is run in opposite directions, producing a spherical quadrupole trapping field. Contours are plotted every 0.35 T for a 1.6 T trap depth. The field is zero at the centre of the cell, and increases in magnitude outwards. Atoms are detected along an axial line passing through the centre of the cell. **b**, Time progression of the Dy spectrum. The panels show spectra of low-field-seeking Dy atoms at increasing times following that

$t = 0$ ablation pulse. The frequency shift is proportional to the magnitude of the magnetic field. The optical density at a given frequency shift is proportional to the number of atoms at the corresponding field. Values shown are means $\pm$ 1 s.e.m. The peak at ~ 30 GHz frequency shift at early times is due to atoms in the high-field, large-volume trap saddle regions shown in **a**. The atoms fall in towards low field (low frequency shift) over about ~ 50 ms.

studied previously, evaporative cooling and buffer-gas loading in a magnetic trap is impossible^{15–21}.

Our experiment demonstrates trapping of non-S-state rare-earth atoms and shows that collisional reorientation of angular momentum in such atoms is highly suppressed. We study the non-S-state rare-earth atoms whose unpaired electrons are exclusively in the 4f shell—Pr, Nd, Tb, Dy, Ho, Er and Tm. The 4f electrons of these atoms are shielded by two outer filled *s* shells (5*s* and 6*s*); the electronic anisotropy of the atoms is submerged under the spherically symmetric *s* shells. As described below, this allows for magnetic trapping of these atoms and opens the possibility for evaporative cooling of all or most of the rare-earth atoms.

Our apparatus is similar to that described in ref. 22. Rare-earth atoms are produced by laser ablation of rare-earth metal foils in a cylindrical copper cell surrounded by a superconducting magnet consisting of two solenoids (Fig. 1). The cell is filled with a ³He buffer gas ($8 \times 10^{15} \text{ cm}^{-3}$) cooled to an initial temperature of 350 mK using a dilution refrigerator. The temperature of the buffer gas rises owing to heating by the ablation pulse and then cools to ~ 800 mK within 20 ms. Rare-earth atoms are detected via absorption spectroscopy by a beam that enters the cell from the bottom window and retro-reflects from a mirror mounted in the centre of the field of the element targets. The rare earths all have strong optical transitions between 410 and 465 nm, so a single laser source is used to probe the atoms. Doing experiments on each element simply requires retuning the laser system and moving the ablation beam to a new target.

We measure both inelastic (Zeeman relaxation) and diffusion (due to elastic scattering) cross-sections for collisions of the rare-earth atoms with ³He. The diffusion cross-section, σ_{el} , for rare-earth/He collisions is determined by measuring the diffusion time of the atoms through the buffer gas at zero magnetic field²³. To measure rare-earth/He inelastic cross-sections, the currents in the two magnet coils are run in the same direction (Helmholtz configuration), producing a strong (up to 2.2 T) homogeneous field. After ablation, the rare-earth atoms translationally thermalize with the buffer gas to a temperature below 2 K within 5 ms, so that the kinetic energy is less than the energy splittings between the Zeeman

levels. Immediately after the ablation, all Zeeman levels are equally populated. Subsequently, atoms in the higher-energy, low-field-seeking Zeeman states relax to lower-energy Zeeman states through inelastic collisions until the populations across Zeeman levels correspond to those of a Boltzmann distribution at the translational temperature of the gas. For each species, the observed decay of atoms in the low-field-seeking state occurs on a timescale ≥ 50 ms, the drift-assisted diffusion time to the cell walls. An upper bound on the inelastic collision cross-section of $\sigma_{\text{in}} < 7 \times 10^{-19} \text{ cm}^2$ is determined from the data. Inelastic collisions are seen to be highly suppressed, with $\gamma \equiv \sigma_{\text{el}}/\sigma_{\text{in}} > 10^4$ in all cases. In order to achieve precooling of low-field-seeking rare-earth atoms while they are in the magnetic trapping fields, the orientation of their magnetic moments with respect to the local magnetic field axis, the atom's Zeeman state, must remain fixed during the approximately 100 elastic collisions required to cool the atom from the initial production temperature of $\sim 1,000$ K to trapping temperatures of ~ 1 K. Furthermore, to achieve a trapped sample, time must be allowed for the low-field-seekers to diffuse towards the centre of the trap. For our trapping geometry, each atom undergoes roughly 30,000 collisions during this time. Thus γ must be $\geq 30,000$ in order to trap using standard buffer-gas loading methods—a constraint satisfied by each of our rare-earth atoms.

To create the magnetic trap, the currents in the solenoids are run in opposite directions (anti-Helmholtz configuration), producing a spherical quadrupole trapping field (Fig. 2a). The trap depths used in this work range from 1 to 3 T. Figure 2b shows early time spectra

Table 1 Summary of trapping results				
Element	Term	Magnetic moment (μ_B)	Number trapped	$\gamma \equiv \sigma_{\text{el}}/\sigma_{\text{in}}$
Tm	² F _{7/2}	3.99	2×10^{11}	$(2.7 \pm 1.4) \times 10^4$
Er	³ H ₆	6.98	2×10^{11}	$(4.3 \pm 2.3) \times 10^4$
Nd	⁵ I ₄	2.41	1×10^{12}	$(8.7 \pm 4.9) \times 10^4$
Tb	⁶ H _{13/2}	9.94	2×10^{11}	$(1.2 \pm 0.6) \times 10^5$
Pr	⁴ I _{9/2}	3.29	3×10^{11}	$(1.3 \pm 0.8) \times 10^5$
Ho	⁴ I _{15/2}	8.96	9×10^{11}	$(2.8 \pm 1.6) \times 10^5$
Dy	⁵ I ₈	9.93	2×10^{12}	$(4.5 \pm 2.4) \times 10^5$

Magnetic moments are given in units of $\mu_B = 1$ Bohr magneton.

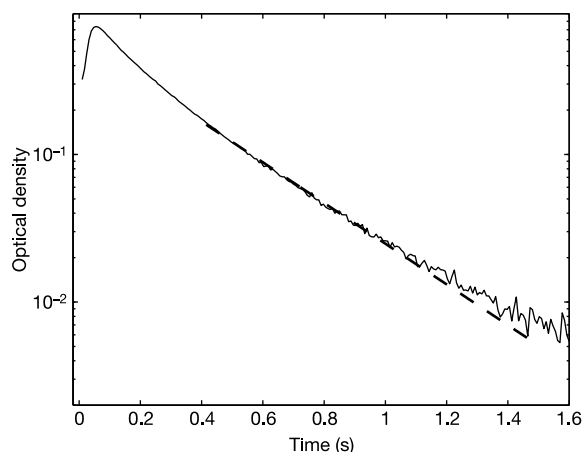


Figure 3 One-body decay of trapped Ho atoms owing to collisionally induced Zeeman relaxation. The decay (solid curve) is fitted (dashed line) to determine a Ho-³He inelastic cross-section of $\sigma_{\text{in}} = (4.7 \pm 2.4) \times 10^{-20} \text{ cm}^2$.

of low-field-seeking Dy atoms following the ablation pulse, obtained via absorption spectroscopy on the $^5\text{I}_8 \rightarrow ^5\text{K}_9$ transition at 421 nm. As the Zeeman level shifts in these atoms are linear with the magnitude of the applied magnetic field, the frequency shift is proportional to the magnetic field (4–15 GHz per T). The optical density at a given frequency shift is proportional to the number of atoms at the corresponding magnetic field strength. The peak near zero frequency shift builds with time, reflecting the movement of low-field-seeking atoms to the lower fields at the centre of the trap. This initial falling-in is complete by ~ 50 ms after ablation. After this time, the spectrum indicates a thermalized trapped atom distribution, and the atoms decay uniformly in a time set by the Zeeman relaxation in collisions with buffer gas helium atoms. The spectra are compared with theoretical simulations of the trapped ensemble to determine the temperature and the number of trapped atoms. We find that, for each species, $(0.2\text{--}2) \times 10^{12}$ rare-earth atoms are trapped at densities of $(0.2\text{--}8) \times 10^{12} \text{ cm}^{-3}$ and temperatures of ~ 800 mK (Table 1).

We determine the rare-earth/He inelastic collision rates by measuring the decay of atoms from the trap. The trap depth (4.5–11 K, depending on the species) is much larger than the temperature of the buffer gas, so we attribute any one-body loss to inelastic collisions with the buffer gas He atoms. Figure 3 shows the optical density versus time at a fixed frequency shift corresponding to low-field-seeking Ho atoms near the centre of the trap. After the initial falling-in of atoms into the centre of the trap, we observe one-body, exponential decay. The values for σ_{in} found from fitting the data for each species (and corresponding values for γ , see Table 1) are compatible with the upper bound deduced from the measurements made in the Helmholtz field.

Following trapping, the next step towards an ultracold gas is to thermally isolate the trapped atoms by removing the buffer gas. We cryopump buffer gas from the trap region by cooling the cell after the ablation pulse, allowing the helium to condense onto the cell walls. Here we use a ⁴He buffer gas to more easily achieve good vacuum after trapping²⁴. Removing the buffer gas takes about 1 s, short enough that we can break thermal contact for the longest lived species, Dy and Ho. The resulting spectrum for Dy is shown in Fig. 4. We confine and thermally isolate $>10^{11}$ Dy atoms and observe the trapped sample for longer than 20 s. The mechanism for decay from the trap is currently under investigation.

As described elsewhere²⁵, it is possible to remove the buffer gas on a much faster (50 ms) timescale by pumping the helium out of the cell through a large-aperture cryogenic valve. This timescale is shorter than all the trap lifetimes we observe, indicating that

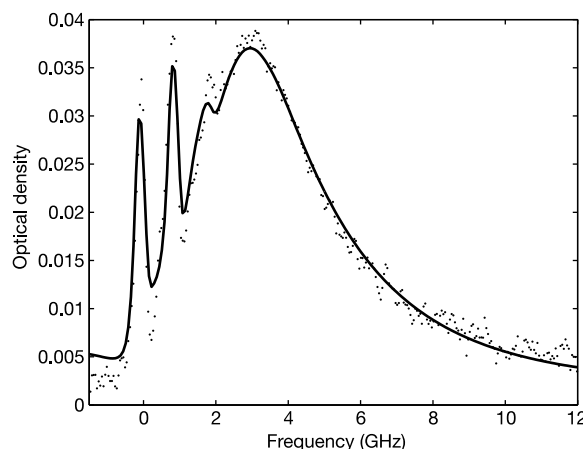


Figure 4 Trapped Dy spectrum following the removal of the ⁴He buffer gas. The measured spectrum is shown by filled circles. The solid curve is a simulation of a trapped thermal distribution with 1×10^{11} Dy atoms trapped at a temperature of 600 mK.

thermal isolation of each trapped rare-earth element would be straightforward by incorporating such a valve into our apparatus.

In conclusion, we have measured the rate for Zeeman relaxation of Pr, Nd, Tb, Dy, Ho, Er and Tm atoms in collisions with helium. All of the atoms we studied were magnetically trapped for the first time. We found that collisional transfer of angular momentum in these atoms is dramatically suppressed, indicating the suppression of electronic interaction anisotropy between rare-earth atoms and He. The ‘submerged shell’ nature of rare-earth atoms is characteristic for about one-third of the elements in the periodic table, indicating that magnetic trapping can be extended to many more species than currently demonstrated.

The trapping of rare-earth atoms may have many applications, ranging from quantum computation with high-magnetic-moment atoms¹³ (for example, Dy) to precision measurements in an effort towards determining the time variation of the fine structure constant^{10–12}. We note that quantum degenerate gases of species with strong dipolar interaction may have unusual, interesting properties, such as anisotropic stability and an altered excitation spectrum. Finally, our work indicates the possibility of creating fermionic and bosonic quantum gases with non-S-state atoms, which may lead to new and interesting physics. □

Received 20 July; accepted 16 August 2004; doi:10.1038/nature02938.

- Wynar, R., Freeland, R. S., Han, D. J., Ryu, C. & Heinzen, D. J. Molecules in a Bose-Einstein condensate. *Science* **287**, 1016–1019 (2002).
- Donley, E. A., Claussen, N. R., Thompson, S. T. & Wieman, C. E. Atom-molecule coherence in a Bose-Einstein condensate. *Nature* **417**, 529–533 (2002).
- Greiner, M., Regal, C. A. & Jin, D. S. Emergence of a molecular Bose-Einstein condensate from a Fermi gas. *Nature* **426**, 537–540 (2003).
- Jochim, S. *et al.* Bose-Einstein condensation of molecules. *Science* **302**, 2101–2103 (2003).
- Widera, A. *et al.* Entanglement interferometry for precision measurement of atomic scattering properties. *Phys. Rev. Lett.* **92**, 160406 (2004).
- Lin, Y.-J., Teper, I., Chin, C. & Vuletic, V. Impact of the Casimir-Polder potential and Johnson noise on Bose-Einstein condensate stability near surfaces. *Phys. Rev. Lett.* **92**, 050404 (2004).
- Barrett, M. D., Sauer, J. A. & Chapman, M. S. All-optical formation of an atomic Bose-Einstein condensate. *Phys. Rev. Lett.* **87**, 010404 (2001).
- Granade, S. R., Gehm, M. E., O'Hara, K. M. & Thomas, J. E. All-optical production of a degenerate Fermi gas. *Phys. Rev. Lett.* **88**, 120405 (2002).
- Takasu, Y. *et al.* Spin-singlet Bose-Einstein condensation of two-electron atoms. *Phys. Rev. Lett.* **91**, 040404 (2003).
- Nguyen, A. T., Budker, D., Lamoreaux, S. K. & Torgerson, J. R. Towards a sensitive search for variation of the fine-structure constant using radio-frequency E1 transitions in atomic dysprosium. *Phys. Rev. A* **69**, 022105 (2004).
- Dzuba, V. A., Flambaum, V. V. & Marchenko, M. V. Relativistic effects in Sr, Dy, YbII, and YbIII and search for variation of the fine-structure constant. *Phys. Rev. A* **68**, 022506 (2003).
- Dzuba, V. A., Flambaum, V. V. & Webb, J. K. Space-time variation of physical constants and relativistic corrections in atoms. *Phys. Rev. Lett.* **82**, 888–891 (1999).
- Derevianko, A. & Cannon, C. C. Quantum computing with magnetically interacting atoms. Preprint at (<http://www.arXiv.org/quant-ph/0406117>) (2004).
- Krems, R. V., Groenenboom, G. C. & Dalgarno, A. Electronic interaction anisotropy between atoms in

- arbitrary angular momentum states. *J. Phys. Chem. A* (in the press).
15. Krems, R. V. & Dalgarno, A. Disalignment transitions in cold collisions of ^3P atoms with structureless targets in a magnetic field. *Phys. Rev. A* **68**, 013406 (2003).
 16. Kokouline, V., Santra, R. & Greene, C. H. Multichannel cold collisions between metastable Sr atoms. *Phys. Rev. Lett.* **90**, 253201 (2003).
 17. Schmidt, P. O. *et al.* Continuous loading of cold atoms into a Ioffe-Pritchard magnetic trap. *J. Opt. B* **5**, S170–S177 (2003).
 18. Xu, X., Loftus, T. H., Hall, J. L., Gallagher, A. & Ye, J. Cooling and trapping of atomic strontium. *J. Opt. Soc. Am. B* **20**, 968–976 (2003).
 19. Nagel, S. B. *et al.* Magnetic trapping of metastable $^3\text{P}_2$ atomic strontium. *Phys. Rev. A* **67**, 011401 (2003).
 20. Katori, H., Ido, T., Isoya, Y. & Kuwata-Gonokami, M. in *Atomic Physics 17* (eds Arimondo, E., DeNatale, P. & Inguscio, M.) 382–396 (AIP, Melville, New York, 2001).
 21. Hansen, D. P., Mohr, J. R. & Hemmerich, A. Magnetic trapping of metastable calcium atoms. *Phys. Rev. A* **67**, 21401 (2003).
 22. Kim, J. *et al.* Buffer-gas loading and magnetic trapping of atomic europium. *Phys. Rev. Lett.* **78**, 3665–3668 (1997).
 23. Hasted, J. B. *Physics of Atomic Collisions* 2nd edn, Ch. 1, 19–45 (American Elsevier, New York, 1972).
 24. deCarvalho, R. *et al.* Buffer-gas loaded magnetic traps for atoms and molecules: A primer. *Eur. Phys. J. D* **7**, 289–309 (1999).
 25. Harris, J. G. E., Michniak, R. A., Nguyen, S. V., Ketterle, W. & Doyle, J. M. Buffer gas cooling and trapping of atoms with small magnetic moments. *Europhys. Lett.* (in the press).

Acknowledgements We thank R. Krems for discussions and for assistance in preparing the manuscript. This material is based on work supported by the National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.M.D. (doyle@physics.harvard.edu).

A tunable carbon nanotube electromechanical oscillator

Vera Sazonova*, Yuval Yaish*, Hande Üstünel, David Roundy, Tomás A. Arias & Paul L. McEuen

Laboratory of Atomic and Solid-State Physics, Cornell University, Ithaca, New York 14853, USA

* These authors contributed equally to this work

Nanoelectromechanical systems (NEMS) hold promise for a number of scientific and technological applications. In particular, NEMS oscillators have been proposed for use in ultrasensitive mass detection^{1,2}, radio-frequency signal processing^{3,4}, and as a model system for exploring quantum phenomena in macroscopic systems^{5,6}. Perhaps the ultimate material for these applications is a carbon nanotube. They are the stiffest material known, have low density, ultrasmall cross-sections and can be defect-free. Equally important, a nanotube can act as a transistor⁷ and thus may be able to sense its own motion. In spite of this great promise, a room-temperature, self-detecting nanotube oscillator has not been realized, although some progress has been made^{8–12}. Here we report the electrical actuation and detection of the guitar-string-like oscillation modes of doubly clamped nanotube oscillators. We show that the resonance frequency can be widely tuned and that the devices can be used to transduce very small forces.

Figure 1a shows a diagram of the measurement geometry and a scanning electron microscope (SEM) image of a device. The fabrication steps have been described elsewhere¹³; briefly, nanotubes (typically single- or few-walled, 1–4 nm in diameter and grown by chemical vapour deposition¹⁴) are suspended over a trench (typically 1.2–1.5 μm wide, 500 nm deep) between two metal (Au/Cr) electrodes. A small section of the tube resides on the oxide on both sides of the trench; the adhesion of the nanotube to the oxide¹⁵ provides clamping at the suspension points.

The measurement is done in a vacuum chamber at pressures below 10^{-4} torr. We actuate and detect the nanotube motion using the electrostatic interaction with the gate electrode underneath the tube. A gate voltage, V_g , induces an additional charge on the nanotube given by $q = C_g V_g$, where C_g is the capacitance between the gate and the tube. The attraction between the charge q and its opposite charge $-q$ on the gate causes an electrostatic force downward on the nanotube. If $C'_g = dC_g/dz$ is the derivative of the gate capacitance with respect to the distance between the tube and the gate, the total electrostatic force on the tube is given by:

$$F_{el} = 1/2 C'_g V_g^2 \cong 1/2 C'_g V_g^{DC} (V_g^{DC} + 2\delta V_g) \quad (1)$$

where we have assumed that the gate voltage has both a static (DC) component and a small time-varying (AC) component. The DC voltage V_g^{DC} at the gate produces a static force on the nanotube that can be used to control its tension. The AC voltage δV_g produces a periodic electric force, which sets the nanotube into motion. As the driving frequency ω approaches the resonance frequency ω_o of the tube, the displacement becomes large.

To detect the vibrational motion of the nanotube, we employ the transistor properties of semiconducting¹⁶ and small-bandgap semiconducting carbon nanotubes^{17,18}, that is, that the conductance change is proportional to the change in the induced charge q on the tube.

$$\delta q = \delta(C_g V_g) = C_g \delta V_g + V_g \delta C_g \quad (2)$$

The first term is the standard transistor gating effect—the modu-

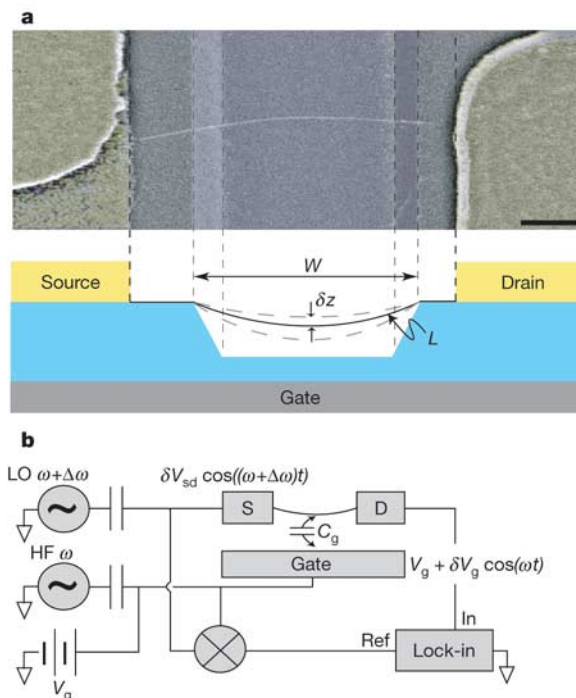


Figure 1 Device geometry and diagram of experimental set-up. **a**, A false-colour SEM image of a suspended device (top) and a schematic of device geometry (bottom). Scale bar, 300 nm. Metal electrodes (Au/Cr) are shown in yellow, and the silicon oxide surface in grey. The sides of the trench, typically 1.2–1.5 μm wide and 500 nm deep, are marked with dashed lines. A suspended nanotube can be seen bridging the trench. CVD growth is known to produce predominantly single- and double-walled nanotubes, but we did not perform detailed studies of the number of walls for the nanotubes on our samples. **b**, A diagram of the experimental set-up. A local oscillator (LO) voltage $\delta V_{sd}^{\omega+\Delta\omega}$ (usually around 7 mV) is applied to the source (S) electrode at a frequency offset from the high frequency (HF) gate voltage signal δV_g^{ω} by an intermediate frequency $\Delta\omega$ of 10 kHz. The current from the nanotube is detected by a lock-in amplifier through the drain electrode (D), at $\Delta\omega$, with time constant of 100 ms.

lation of conductance due to the modulation of the gate at the driving frequency—and it is observed at any driving frequency. The second term is non-zero only if the tube moves (when the driving frequency approaches the resonance); the distance to the gate changes, resulting in a variation δC_g in its capacitance.

To detect this conductance change we use the nanotube as a mixer¹⁹ (Fig. 1b). This method helps avoid unnecessary complications due to capacitive currents between the gate and the drain electrodes. The magnitude of the current is given by the product of the AC voltage on the source electrode δV_{sd} , and the modulated nanotube conductance δG . Using equation (2) we derive the result that the expected current is:

$$\delta I^{\text{lock-in}} = \delta G \delta V_{sd} = \frac{1}{2\sqrt{2}} \frac{dG}{dV_g} \left(\delta V_g + V_g^{\text{DC}} \frac{\delta C_g}{C_g} \right) \delta V_{sd} \quad (3)$$

where δV_g is the AC voltage applied to the gate electrode.

Figure 2a shows the measured current as a function of driving frequency at room temperature. We see a distinctive feature in the current on top of a slowly changing background. We attribute this feature to the resonant motion of the nanotube, modulating the capacitance, while the background is due to modulating gate voltage. The response fits well to a lorentzian function with a normalized linewidth $Q^{-1} = \Delta f/f_o = 1/80$, a resonant frequency $f_o = 55$ MHz, and an appropriate phase difference between the actuation voltage and the force on the nanotube¹⁹.

The DC voltage on the gate can be used to tune the tension in the nanotube and therefore the oscillation frequency. Figure 2b and c show colour-scale plots of the measured response as a function of the driving frequency and the static gate voltage. The resonant

frequency shifts upward as the magnitude of the DC gate voltage is increased. Several distinct resonances are observed, corresponding to different vibrational modes of the nanotube. We have found similar results in 11 comparable devices, with resonance frequencies varying from 3 to 200 MHz for different samples and gate voltages.

To understand the frequency dependence of the nanotube oscillations with V_g^{DC} , we have performed a series of simulations of the vibrational properties of nanotubes. We model the nanotube as a slack beam suspended over a trench. Slack here means that the tube is longer than the distance between the contacts, which is a result of the nanotube's curvature before suspension. Slack was observed for almost all imaged devices in an SEM (Fig. 1a), and has also been inferred from atomic force microscope (AFM) force measurements of similar samples¹³. A finite element model is then used to calculate the vibrational frequencies for a nanotube with a typical geometry (length $L = 1.75 \mu\text{m}$, radius $r = 1$ nm) and mechanical rigidities determined using the Tersoff–Brenner potential²⁰.

The theoretical results for a representative device can be seen in Fig. 2d. For no static electric force on the nanotube, $V_g^{\text{DC}} \approx 0$, the resonance frequency is determined by the bending rigidity of the nanotube and is approximately that of an equivalent doubly clamped beam with no tension. At small V_g^{DC} , there is a static electric force downward on the nanotube (equation (1)), producing a tension $T \propto V_g^2$, which shifts the resonant frequency of the nanotube, $\Delta\omega_0 \propto T \propto V_g^2$ (ref. 21). At intermediate V_g^{DC} , the electrostatic force overcomes the bending rigidity and the nanotube behaves as a hanging chain; the profile of the tube forms a catenary. In this regime, the resonance frequency is given by $\omega_0 \propto \sqrt{T} \propto V_g$. In the large electrostatic force regime, the nanotube behaves as an

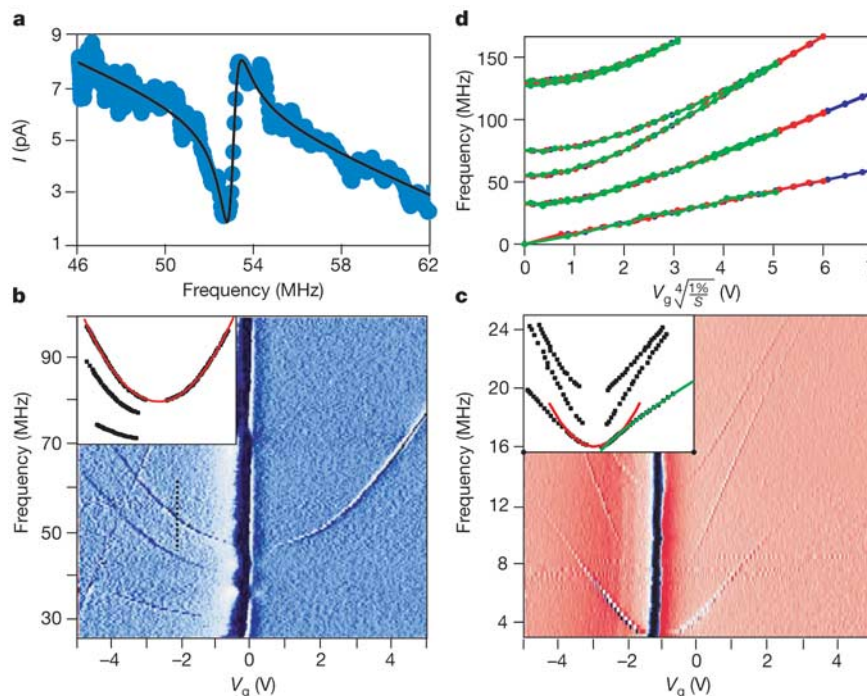


Figure 2 Measurements of the resonant response. The measurements were done on 11 devices, both semiconducting and small bandgap semiconducting nanotubes in a vacuum chamber at pressures below 10^{-4} torr. The maximum conductance G^{max} , and the transconductance dG/dV_g^{max} , are given below for the presented devices. **a**, Detected current as a function of driving frequency taken at $V_g = 2.2$ V, $\delta V_g = 7$ mV for device 1 ($G^{\text{max}} = 12.5 \mu\text{S}$, $dG/dV_g^{\text{max}} = 7 \mu\text{S V}^{-1}$). The solid black line is a lorentzian fit to the data with an appropriate phase shift between the driving voltage and the oscillation of the tube. The fit yields the resonant frequency $f_o = 55$ MHz, and quality factor $Q = 80$. **b, c**, Detected current (plotted as a derivative in colour scale) as a function of gate voltage and frequency for devices 1 and 2 ($G^{\text{max}} = 10 \mu\text{S}$, $dG/dV_g^{\text{max}} = 0.3 \mu\text{S V}^{-1}$). Panel **a** is

a vertical slice through panel **b** at $V_g = 2.2$ V (marked with a dashed black line). The insets to the figures show the extracted positions of the peaks in the frequency–gate voltage space for the respective colour plots. A parabolic and a $V_g^{2/3}$ fit of the peak position are shown in red and green, respectively. **d**, Theoretical predictions for the dependence of vibration frequency on gate voltage for a typical device with length $L = 1.75 \mu\text{m}$, and radius $r = 1$ nm. The calculations were performed for several different values of slack s ($s = (L - W)/W$, where L is the tube's length and W is the distance between clamping points). The calculations for 0.5%, 1% and 2% slack are shown in blue, red and green, respectively. Notice the appropriately rescaled x-axis.

elastic string—the extensional rigidity becomes dominant. In this regime, the resonance frequency is given by $\omega_0 \propto \sqrt{T} \propto V_g^{2/3}$ (ref. 21). The transition point from the bending-dominated regime to the catenary, and from the catenary to the stretching-dominated regime, depends on the amount of slack in the nanotube. Either one, two or all three described regimes may be relevant for a particular device. For the bending-dominated and catenary regimes, the voltage dependence of the resonant frequencies scales as the slack to the 1/4th power (Fig. 2d).

Comparing these predictions with Fig. 2b and c, we see a good qualitative agreement with predicted dispersions. All of the resonances start dispersing parabolically, some continuing into a linear

regime as the gate voltage is increased. For the lowest resonance shown in Fig. 2c we can also observe the $\omega_0 \propto V_g^{2/3}$ frequency dependence at large gate voltages. The frequency dependence of the resonances is thus in good qualitative agreement with theoretical expectations. We do, however, often find multiple resonances lower in frequency than is predicted by the theoretical calculations. One low-frequency mode is expected because any small asymmetric clamping will result in a non-zero frequency at zero gate voltage for the lowest branch in Fig. 2d. The additional low-frequency modes could be caused by extra mass due to contaminants coating the nanotube, or by a large asymmetry in the clamping conditions. Further studies are needed to understand the exact nature of this frequency lowering.

To determine the other parameters of the nanotube oscillator, we have studied the dependence of the measured resonance on the amplitude of the gate drive signal δV_g . Figure 3a shows results for one device. For low driving amplitudes, the response on resonance is linear in δV_g and Q is roughly constant. As the δV_g is increased further, the response saturates and the quality factor decreases. For some devices, there is also a dramatic change in the signal shape observed at these high driving voltages (Fig. 3b). Instead of a smooth lorentzian dip, the system develops a hysteretic transition between low- and high-amplitude states of oscillation.

To understand these results, we first address the linear response regime. We estimate the amplitude δz of the nanotube oscillation using the measured signal amplitude relative to the background in conjunction with equation (3). From this, we can extract the relative change in the capacitance $\delta C_g/C_g$ on resonance; for the data in Fig. 2a, where $\delta V_g = 7$ mV, we obtain $\delta C_g/C_g = 0.3\%$. Assuming a logarithmic model for capacitance $C_g = \frac{4\pi\epsilon_0 L}{2\ln(2z/r)}$, where L is the suspended length of the NT, and z is the distance to the gate, this can be translated into a distance change, $\frac{\delta z}{z} = \frac{\delta C_g}{C_g} \ln(2z/r)$. In Fig. 2a, we estimate the amplitude of motion to be $\delta z \approx 10$ nm. Calculating the driving force using equation (1), we get $F = C'_g V_g^{DC} \delta V_g \approx 60$ fN. Thus, we estimate the effective spring constant for this resonance to be $k_{\text{eff}} = \frac{F}{\delta z} Q \approx 4 \times 10^{-4} \text{ N m}^{-1}$. Note that this effective spring constant is different for each resonance.

As the amplitude of the oscillation is increased, we can expect the nonlinear effects due to the change in spring constant to become important. It is well known that nonlinear oscillators have a bistable region in their response-frequency phase space which experimentally results in a hysteretic response²². The onset of nonlinear effects in our case corresponds to driving voltages of 15 mV. Assuming the same parameters as above yields an amplitude of motion of 30 nm.

Another important parameter characterizing the oscillator is the quality factor Q , the ratio of the energy stored in the oscillator to the energy lost per cycle owing to damping. Maximizing Q is important for most applications. It is in the range of 40–200 for our samples with no observed frequency dependence. Previous measurements on larger multiwalled nanotubes at room temperature and ropes of single-walled nanotubes at low temperatures yielded^{8–11} values of Q in the range 150–2,500.

Because one source of dissipation could be air drag, we have studied the dependence of the resonator properties on the pressure in the vacuum chamber. Figure 3c summarizes the results for one device. Q decreases with pressure, and the resonance is no longer observed above pressures of 10 torr. This is in good agreement with calculations²³. At lower pressures, air losses should be minimal. Many other sources could be contributing to the damping, including the motion of surface adsorbates and ohmic losses due to the motion of electrons on and off the tube. The former is difficult to estimate, but we have calculated the magnitude of the latter and find it to be insignificant. Another important potential source of dissipation is clamping losses where the nanotube is attached to the substrate; the tube may lose energy by sticking and unsticking from the surface during oscillation. Experiments on devices with different clamping geometries are necessary to investigate this issue.

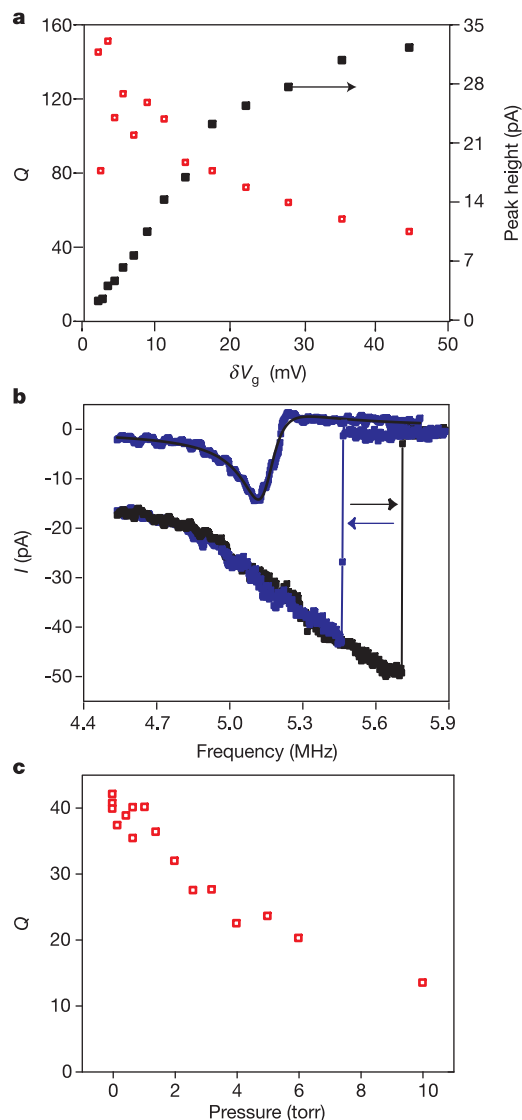


Figure 3 Amplitude and pressure dependence of the resonance. **a** The measured quality factor Q of the resonance and the height of the resonance peak for device 3 ($G^{\text{max}} = 15 \mu\text{S}$, $dG/dV_g^{\text{max}} = 4 \mu\text{S V}^{-1}$) are shown in red open squares and black solid squares, respectively, as a function of driving voltage δV_g . Linear behaviour is observed at low voltages, but Q decreases and the height of the peak saturates at higher driving voltages. **b**, Trace of detected current versus frequency with the background signal subtracted for device 2 at two different driving voltages $\delta V_g = 8.8$ mV and $\delta V_g = 40$ mV. The solid black line is a Lorentzian fit to the low bias data. The traces of the current as the frequency is swept up and down are shown in blue and black, respectively. Hysteretic switching can be observed. **c**, Pressure dependence of the resonance peak for device 4 ($G^{\text{max}} = 7.7 \mu\text{S}$, $dG/dV_g^{\text{max}} = 0.6 \mu\text{S V}^{-1}$). The Q of the resonance peak is shown in red open squares. The peak was no longer observed above pressures of 10 torr.

The nanotube oscillator parameters presented above are representative of all of our measured devices. Using these parameters, we can calculate the force sensitivity of the device at room temperature. The smallest detected motion of the nanotube was at a resonant driving voltage of $\delta V_g \approx 1$ mV in the bandwidth of 10 Hz. The sensitivity was limited by the Johnson–Nyquist electronic noise from the nanotube. Using equations (1) and (3) above, this corresponds to a motion of ~ 0.5 nm on resonance and a force sensitivity of ~ 1 fN Hz $^{-1/2}$. This is within a factor of ten of the highest force sensitivities measured at room temperature²⁴.

The ultimate limit on force sensitivity is set by the thermal vibrations of the nanotube. The corresponding force sensitivity is $\delta F_{\min} = \sqrt{\frac{4k_B kT}{\omega_0 Q}} = 20$ aN Hz $^{-1/2}$ for typical parameters. The observed sensitivity is 50 times lower than this limit. This is probably due to the relatively low values of transconductance for the measured nanotubes at room temperature. At low temperatures (~ 1 K), the sensitivity should increase by orders of magnitude owing to the high transconductance associated with Coulomb oscillations¹⁹. Even without increasing Q , force sensitivities below 5 aN should theoretically be attainable at low temperatures. This is comparable to the highest sensitivities measured^{25–28}. The combination of high sensitivity, tunability, and high-frequency operation make nanotube oscillators promising for a variety of scientific and technological applications. □

Received 29 April; accepted 29 July 2004; doi:10.1038/nature02905.

- Sidles, J. A. *et al.* Magnetic-resonance force microscopy. *Rev. Mod. Phys.* **67**, 249–265 (1995).
- Roukes, M. Plenty of room indeed. *Sci. Am.* **285**, 48–57 (2001).
- Nguyen, C. T. C. Frequency-selective MEMS for miniaturized low-power communication devices. *IEEE Trans. Microwave Theory Tech.* **47**, 1486–1503 (1999).
- Nguyen, C. T. C., Wong, A.-C. & Hao, D. *Solid-State Circuits Conf.* 78–79, 1999 (Digest of Technical Papers, ISSCC, IEEE International, San Francisco, California, 1999).
- Cho, A. Physics—Researchers race to put the quantum into mechanics. *Science* **299**, 36–37 (2003).
- LaHaye, M. D., Buu, O., Camarota, B. & Schwab, K. C. Approaching the quantum limit of a nanomechanical resonator. *Science* **304**, 74–77 (2004).
- Dresselhaus, M. S., Dresselhaus, G. & Avouris, P. *Carbon Nanotubes* (Springer, Berlin, 2001).
- Poncharal, P., Wang, Z. L., Ugarte, D. & de Heer, W. A. Electrostatic deflections and electromechanical resonances of carbon nanotubes. *Science* **283**, 1513–1516 (1999).
- Gao, R. P. *et al.* Nanomechanics of individual carbon nanotubes from pyrolytically grown arrays. *Phys. Rev. Lett.* **85**, 622–625 (2000).
- Reulet, B. *et al.* Acoustoelectric effects in carbon nanotubes. *Phys. Rev. Lett.* **85**, 2829–2832 (2000).
- Purcell, S. T., Vincent, P., Journet, C. & Binh, V. T. Tuning of nanotube mechanical resonances by electric field pulling. *Phys. Rev. Lett.* **89**, 276103 (2002).
- Babic, B., Furer, J., Sahoo, S., Farhangfar, S. & Schonenberger, C. Intrinsic thermal vibrations of suspended doubly clamped single-wall carbon nanotubes. *Nano Lett.* **3**, 1577–1580 (2003).
- Minot, E. D. *et al.* Tuning carbon nanotube band gaps with strain. *Phys. Rev. Lett.* **90**, 156401 (2003).
- Kong, J., Soh, H. T., Cassell, A. M., Quate, C. F. & Dai, H. J. Synthesis of individual single-walled carbon nanotubes on patterned silicon wafers. *Nature* **395**, 878–881 (1998).
- Hertel, T., Walkup, R. E. & Avouris, P. Deformation of carbon nanotubes by surface van der Waals forces. *Phys. Rev. B* **58**, 13870–13873 (1998).
- Tans, S. J., Verschueren, A. R. M. & Dekker, C. Room-temperature transistor based on a single carbon nanotube. *Nature* **393**, 49–52 (1998).
- Zhou, C. W., Kong, J. & Dai, H. J. Intrinsic electrical properties of individual single-walled carbon nanotubes with small band gaps. *Phys. Rev. Lett.* **84**, 5604–5607 (2000).
- Minot, E. D., Yaish, Y., Sazonova, V. & McEuen, P. L. Determination of electron orbital magnetic moments in carbon nanotubes. *Nature* **428**, 536–539 (2004).
- Knobel, R. G. & Cleland, A. N. Nanometre-scale displacement sensing using a single electron transistor. *Nature* **424**, 291–293 (2003).
- Brenner, D. W. Empirical potential for hydrocarbons for use in simulating the chemical vapor-deposition of diamond films. *Phys. Rev. B* **42**, 9458–9471 (1990).
- Sapmaz, S., Blanter, Y. M., Gurevich, L. & van der Zant, H. S. J. Carbon nanotubes as nanoelectromechanical systems. *Phys. Rev. B* **67**, 235414 (2003).
- Yurke, B., Greywall, D. S., Pargellis, A. N. & Busch, P. A. Theory of amplifier-noise evasion in an oscillator employing a nonlinear resonator. *Phys. Rev. A* **51**, 4211–4229 (1995).
- Bhildvala, R. B. & Wang, Z. J. Effect of fluids on the Q-factors and resonance frequency of oscillating beams. *Phys. Rev. E* **69**, 036307 (2004).
- Jenkins, N. E. *et al.* Batch fabrication and characterization of ultrasensitive cantilevers with submicron magnetic tips. *J. Vac. Sci. Technol. B* **22**, 909–915 (2004).
- Stowe, T. D. *et al.* Attonewton force detection using ultrathin silicon cantilevers. *Appl. Phys. Lett.* **71**, 288–290 (1997).
- Mohanty, P., Harrington, D. A. & Roukes, M. L. Measurement of small forces in micron-sized resonators. *Physica B* **284**, 2143–2144 (2000).
- Stipe, B. C., Mamin, H. J., Stowe, T. D., Kenny, T. W. & Rugar, D. Magnetic dissipation and fluctuations in individual nanomagnets measured by ultrasensitive cantilever magnetometry. *Phys. Rev. Lett.* **86**, 2874–2877 (2001).
- Mamin, H. J. & Rugar, D. Sub-attonewton force detection at millikelvin temperatures. *Appl. Phys. Lett.* **79**, 3358–3360 (2001).

Acknowledgements We thank E. Minot for discussions. This work was supported by the NSF through the Cornell Center for Materials Research and the NIRT program, and by the MARCO Focused Research Center on Materials, Structures, and Devices. Sample fabrication was performed at the Cornell Nano-Scale Science and Technology Facility (a member of the National Nanofabrication Infrastructure Network), funded by the NSF.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.L.M. (mceuen@ccmr.cornell.edu).

Supramolecular self-assembled molecules as organic directing agent for synthesis of zeolites

Avelino Corma¹, Fernando Rey¹, Jordi Rius², Maria J. Sabater¹ & Susana Valencia¹

¹Instituto de Tecnología Química, UPV-CSIC, Universidad Politécnica de Valencia, Avda de los Naranjos s/n, 46022 Valencia, Spain

²Institut de Ciència de Materials de Barcelona (CSIC), Campus de la UAB, 08193 Bellaterra, Catalunya, Spain

Solid materials with uniform micropores, such as zeolites, can act as selective catalysts and adsorbents for molecular mixtures by separating those molecules small enough to enter their pores while leaving the larger molecules behind^{1,2}. Zeolite A is a microporous material with a high void volume. Despite its widespread industrial use in, for example, molecular separations and in detergency^{3,4}, its capability as a petroleum-refining material is limited owing to its poor acid-catalytic activity and hydrothermal stability, and its low hydrophobicity. These characteristics are ultimately a consequence of the low framework Si/Al ratio (normally around one) and the resulting high cationic fraction within the pores and cavities^{1,2}. Researchers have modified the properties of type-A zeolites by increasing the Si/Al compositions up to a ratio of three^{5–9}. Here we describe the synthesis of zeolite A structures exhibiting high Si/Al ratios up to infinity (pure silica). We synthesize these materials, named ITQ-29, using a supramolecular organic structure-directing agent obtained by the self-assembly, through π – π type interactions, of two identical organic cationic moieties. The highly hydrophobic pure-silica zeolite A can be used for hydrocarbon separations that avoid oligomerization reactions, whereas materials with high Si/Al ratios give excellent shape-selective cracking additives for increasing propylene yield in fluid catalytic cracking operations. We have also extended the use of our supramolecular structure-directing agents to the synthesis of a range of other zeolites.

Zeolite A has an ‘LTA’ structure¹⁰, which can be envisaged as a three-dimensional network of spherical ~ 1.14 nm cavities (α -cages), interconnected by six small windows that are limited by eight-membered ring (8MR) pores with a diameter of 0.41 nm (see Fig. 1). This structure can also be described in terms of sodalite cages (β -cages), which are connected through the square faces with each other, forming double four-ring units (D4R). Typically, it is synthesized in the sodium form with framework Si/Al ratios of about one, the composition of the unit cell being: $[\text{Na}_{12}\text{Al}_{12}\text{Si}_{12}\text{O}_{48}] \cdot (\text{H}_2\text{O})_{27}$. The adsorption capacity of this zeolite can be increased by replacing the sodium with calcium cations in a post-synthesis exchange. The calcium form of zeolite A (CaA) already shows good adsorption selectivity that allows the normal

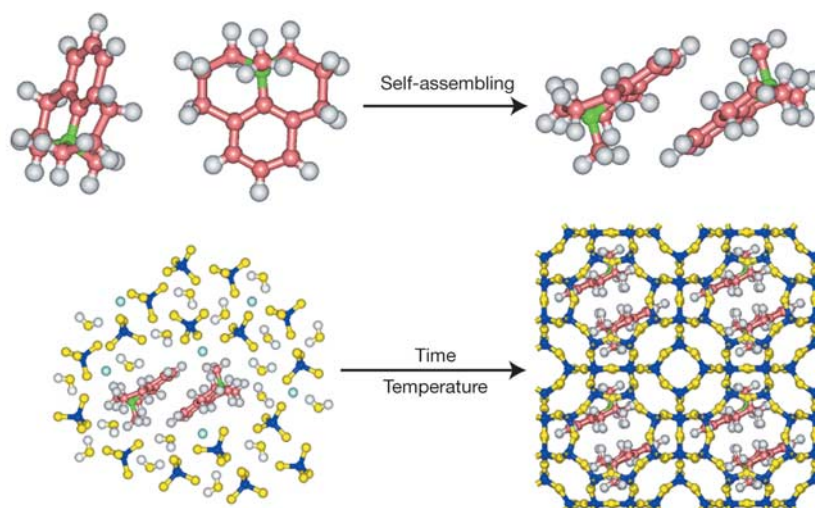


Figure 1 Formation of the LTA structure from the supramolecular self-assembling of the OSDA molecules. See Supplementary Data. Atoms are colour-coded as follows: dark blue, silicon; light blue, fluorine; yellow, oxygen; red, carbon; green: nitrogen; grey, hydrogen.

short-chain alkanes to be separated from the branched ones, because the size of the 8MR windows gives access to the cavities. However, because of the polarity of CaA, this zeolite cannot be used for olefin/paraffin separations, which require a pure-silica A zeolite. Furthermore, CaA has a limited hydrothermal and acid stability that would be much improved if this zeolite could be prepared with high Si/Al ratios.

It should be possible to extend the adsorption sieving effect of zeolite A to shape selective catalysis if we could introduce acid sites into the structure. Unfortunately, a material with an Si/Al ratio of ~ 1 is not stable enough to support acidity.

It is not surprising that many researchers, especially those from industry, have attempted to synthesize materials having an LTA structure with higher Si/Al ratios. A partial success was achieved by researchers from Union Carbide and Mobil^{5–8}, who obtained zeolites with LTA structure and an Si/Al ratio up to about three by using tetramethylammonium cations (TMA^+), together with sodium ions. They realized that by filling the void volume of the zeolite with large TMA^+ cations, a limitation on the number of tetrahedral aluminium ions in the lattice should occur. Interestingly, the sodium form of those samples, despite its still low Si/Al ratio, showed improved adsorption properties. Furthermore, zeolite A with an Si/Al ratio of three already presented some possibilities for the selective hydrogenation of olefins and hydrocracking of alkanes⁹, but its stability and high Al content was still not sufficient to allow some selective adsorptions and many potential catalytic applications.

To increase further the framework Si/Al ratio, it would be necessary to find bulky organic structure-directing agents (OSDA) that can fit within the zeolite cavity. A successful OSDA should have only a weak tendency to form complexes with the solvent^{11,12}. Moreover, it should fit the inner surface of the cage of the LTA structure with as many van der Waals contacts as possible, but with the least deformation. Finally, the OSDA has to maintain adequate hydrophobicity. It is more difficult to achieve all these requirements simultaneously when a large organic structure-directing agent has to be prepared that will fit within an already pre-selected cavity shape. In the case of zeolite A, finding an adequate OSDA was made more difficult by the fact that the spherical shape of the cavity is delimited by small 8MR pores, and the requirement of a relatively rigid organic molecule would involve clathrasils as competing phases. Furthermore, if the hydrophobicity is inadequate and the organic–inorganic interactions are not optimal, then the structure-

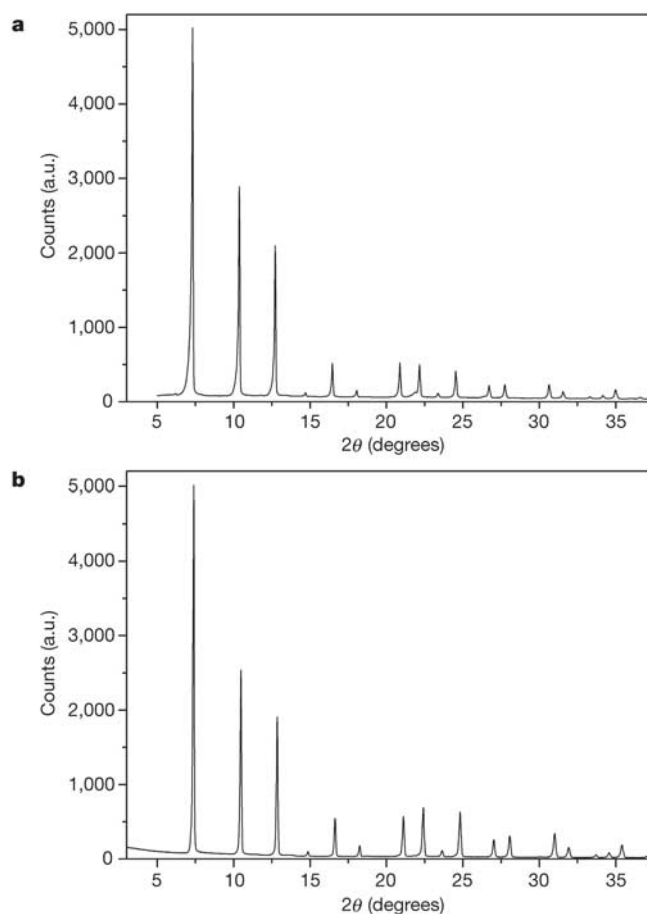


Figure 2 X-ray diffraction patterns of calcined ITQ-29 materials. **a**, Ge-containing zeolite (Si/Ge = 2). **b**, Pure-silica zeolite.

directing effects of the organic will not be determinant, and amorphous material could be the competing phase.

To overcome the above limitations we used the following strategy. From our previous work^{13–16}, we learned that Ge, as well as F^- , can direct the synthesis towards the formation of D4R-containing structures. In fact, ITQ-21, with its six 12MR circular windows,

Table 1 Adsorption properties of ITQ-29 compared with other zeolites

Sample	298 K	Propane (mg g ⁻¹) 313 K	333 K	298 K	Propene (mg g ⁻¹) 313 K	323 K	333 K	Butane (mg g ⁻¹) 353 K	Butene (mg g ⁻¹) 353 K	Hexane (mg g ⁻¹) 353 K	1-Hexene (mg g ⁻¹) 333 K	3-Methylpentene (mg g ⁻¹) 298 K	H ₂ O (mg g ⁻¹) 298 K	Ar (cm ³ g ⁻¹) 87 K
ITQ-29	95	88	75	105	92	88	83	86	105	88	154	1	10	0.24
Ca-A	—	132 (ref. 23; 323 K)	—	—	—	142 (ref. 23)	—	87 (ref. 1; 348 K)	—	122 (ref. 1; 323 K)	—	—	260	0.22
ITQ-3	—	—	30 (ref. 21; 353 K)	—	61	—	53	—	—	—	—	—	—	0.23
ITQ-12	17 (ref. 22; 303 K)	—	—	55 (ref. 22; 303 K)	—	—	—	—	—	—	—	—	—	0.13

Temperatures at which the measurement was made are shown in parentheses.

Table 2 Catalytic properties of ITQ-29 compared with other zeolites

Catalyst	T(IV)/Al	Crystal size (μm)	Cracking kinetic constant (10 ⁻² s ⁻¹)		Methanol to olefins	
			1-hexene	4-methyl-1-pentene	Conversion (%)	Ethylene/propylene
ITQ-29	20	0.6	13.0	5.4	42.1	1.0
ZSM-5	40	0.5	21.4	15.1	—	—
ITQ-3	30	0.5 × 1.5	1.8	1.8	40.4	0.5
ITQ-12	180	—	0.4	0.4	0	—

T(IV)/Al, ratio of tetrahedral atom to Al in catalyst.

resembles zeolite A except that the cages of ITQ-21 are not sodalite cages. Thus, at least initially, we decided to introduce Ge into the gel and to work in F⁻ media. This should help to link the sodalite cages, forming the corresponding D4R units that will direct the synthesis towards the LTA structure instead of the sodalite (see Fig. 1).

For the OSDA, we wanted to maximize the size to charge ratio of the organic ammonium cation, trying to reach the limit for a pure T(IV) LTA, where T(IV) indicates the element in the tetrahedral lattice with oxidation number 4. We were not able to envisage a bulky molecule that would adapt to the cavity while fulfilling the requirements for a successful template described above. We therefore started with the concept that a very large OSDA can be obtained by supramolecular self-assembling of two moieties, each with the proper rigidity and polarity properties. Of course, we could not make use of hydrogen-bonding interactions to achieve the supramolecular self-assembling because the synthesis is carried out in aqueous media. However, perhaps π - π -type interactions could be used to assemble the two starting organic molecules. If so, such an assembly method may have another benefit: by disassembling the two molecules when the synthesis is finished, the OSDA could be recovered, if the pore dimensions allow them to diffuse out¹⁷. Recovery, however, will not be possible in the case of the LTA structure having the large cavity connected by small pores.

Among different organic molecules considered, one of them, the 4-methyl-2,3,6,7-tetrahydro-1H,5H-pyrido [3.2.1-ij] quinolinium iodide (Fig. 1) was able to self-assemble, forming dimers—as deduced from its structural elucidation using single crystal X-ray diffraction (see Supplementary Information part 1). The ability of this OSDA to self-assemble was further studied with photoluminescence spectroscopy. We found that the main emission band in the spectrum of the aqueous solution of the OSDA in the hydroxide form used for the synthesis of ITQ-29 zeolites appears at ~450 nm (its spectrum is nearly identical to that of the solid iodide salt). The presence of this band has been observed previously when π stacking of aromatic rings in polymers occurs¹⁸. These results clearly indicate that the OSDA is forming self-assembled dimers in aqueous solution. Also, the same emission band was detected in the synthesized ITQ-29 zeolite (see Supplementary Information part 2). With this OSDA, Al-free ITQ-29 was synthesized (yield > 90%) with a Si/Ge = 2 (synthesis details are given in Supplementary Information part 3), and the powder X-ray diffraction (XRD) pattern of the calcined sample is given in Fig. 2a. From this, the

LTA crystalline structure of ITQ-29 was confirmed. The XRD pattern was indexed according to a cubic unit cell with refined cell parameter equal to 1.20157(4) nm, where parentheses indicate uncertainty in the last digit. For the Rietveld refinement, the LTA zeolite structure type with *Pm3m* space group symmetry was employed. According to the refined scattering powder at the T site and assuming that it is filled, the unit-cell composition of the calcined sample is Si_{16.6}Ge_{7.4}O₄₈; that is, the refined Si/Ge ratio is 2.2 (Rietveld refinement details are given in Supplementary Information part 4).

The chemical analysis, ¹³C-cross polarization-magic angle spinning-nuclear magnetic resonance (¹³C-CP-MAS-NMR) and photoluminescence spectroscopies of the sample confirm that there are two intact molecules of OSDA, which are forming self-assembled dimers, within the α -cage. The charges associated with the two molecules are neutralized by two F⁻ ions, which are located in the D4R—as evidenced by ¹⁹F MAS-NMR spectroscopy (resonances at -8 and -20 p.p.m.)^{19,20}.

With the synthesis procedure described above, it was possible to introduce Al (T(IV)/T(III) = 80), and ²⁷Al MAS-NMR shows that it is tetrahedrally coordinated (signal at 52 p.p.m.), suggesting that it has been incorporated in framework positions. To further decrease the framework Si/Al ratio, TMA⁺ were also incorporated in the synthesis gel to compensate the framework charges generated by the introduction of Al. In this way, ITQ-29 with a T(IV)/T(III) ratio of 13 and a Si/Ge ratio of two has been synthesized, in which the Al is tetrahedrally coordinated (details are given in Supplementary Information part 3).

We emphasize that by using both TMA⁺ and the quinolinium derivative OSDA it is also possible to synthesize pure-silica ITQ-29 (see Supplementary Information part 3 for detailed conditions). The powder-XRD pattern of the calcined sample is given in Fig. 2b, and from this the cubic LTA structure was confirmed, with a parameter of 1.18671(4) nm (for Rietveld refinement, see Supplementary Information part 4).

The pure-silica ITQ-29 sample is very stable and the structure is maintained after calcination at 900 °C in the presence of atmospheric moisture. The sample is also stable when treated in an aqueous solution of HCl at pH = 1. By following this approach, ITQ-29 can also be synthesized as aluminosilicate (details are given in Supplementary Information part 3).

We have measured the adsorption capacity of the pure-silica ITQ-

29 described here, the results are given in Table 1 and compared with those reported for ITQ-3 and ITQ-12 and CaA zeolites^{1,21–23}. The water adsorption isotherms on CaA and pure-silica ITQ-29 at 20 mbar give 26 wt% and 1 wt% H₂O adsorption capacity, respectively, indicating the hydrophobic character of this pure-silica material. As we expected, this zeolite allows us to separate normal from branched paraffins, as can be seen in Table 1. The adsorption capacities are CaA > ITQ-29 > ITQ-3 ≫ ITQ-12. For olefin adsorption, it should be taken into account that olefin oligomerization occurs on CaA sample at the adsorption temperature²⁴, making this zeolite unsuitable for adsorbing olefins. In contrast, no oligomerization was detected on ITQ-29 and the adsorption capacity remains constant after 20 adsorption–desorption cycles.

Al-containing ITQ-29 zeolites show acid hydroxyl groups at 3,620 and 3,550 cm^{−1} in the infrared spectra. They are catalytically active and the results (Table 2) show a larger diffusion shape selectivity for cracking of 1-hexene and 4-methyl-1-pentene than ZSM-5, which is the commercial catalyst used for the selective cracking of lineal olefins in FCC, and for hydrocracking (dewaxing) of alkanes. Also, the cracking activity of the Al-containing ITQ-29 sample was much higher than that of other small-pore zeolites such as ITQ-3 and ITQ-12. For conversion of methanol to olefins, both ITQ-29 and ITQ-3 are active and selective catalysts, but the second zeolite shows a deactivation rate twice that of ITQ-29. In contrast, ITQ-12 presents a negligible activity owing to the impossibility of preparing this zeolite with a T(IV)/Al ratio lower than 180.

Thus we have shown that by using OSDA in a new way that involves the supramolecular self-assembling of two organic moieties, it has been possible to synthesize the Al-free as well as the pure-silica zeolites with the LTA structure. The latter is a very stable and hydrophobic material and could be used for gas separation in the presence of H₂O or other polar molecules. By introducing Al in framework positions we have been able to produce acid catalysts that give excellent shape selectivity for cracking (preferably lineal rather than branched) olefins. Such catalysts may be useful as an additive for increasing propylene yield in fluid catalytic cracking operations. Moreover, ITQ-29 may be an interesting catalyst for converting methanol into olefins. □

Methods

Synthesis of ITQ-29 zeolites

ITQ-29 zeolites were synthesized in fluoride media from gels of the following general molar compositions: (1 − x)SiO₂:xGeO₂:yAl₂O₃:0.5ROH:0.5HF:zH₂O (1 − x)SiO₂:xGeO₂:yAl₂O₃:0.25ROH:0.25TMAOH:0.5HF:zH₂O where x was varied from 0 to 0.33, y from 0 to 0.07 and z from 2 to 7.

The gels were prepared by hydrolysing tetraethylorthosilicate (TEOS) and aluminium isopropoxide in an aqueous solution of 4-methyl-2,3,6,7-tetrahydro-1H,5H-pyrido[3.2.1-ij] quinolinium hydroxide (ROH) and tetramethylammonium hydroxide (TMAOH), then the appropriate amount of GeO₂ was added and the mixture was stirred until the ethanol formed upon hydrolysis of TEOS and the appropriate excess of water were evaporated to reach the gel composition given above. Finally, an aqueous solution of HF (50%) was added and the mixture was introduced in a Teflon-lined stainless autoclave and heated at 408 or 423 K for times of between 2 and 7 days. Then, the autoclave was cooled down, and the mixture was filtered, washed with distilled water and dried at 100 °C. The preparation of the of 4-methyl-2,3,6,7-tetrahydro-1H,5H-pyrido[3.2.1-ij] quinolinium hydroxide used as OSDA is given in Supplementary Information part 1 and the detailed synthesis procedure and the chemical compositions of the materials are given in Supplementary Information part 2.

The adsorption properties of the ITQ-29 materials were determined by Ar adsorption using a ASAP 2010 instrument from Micromeritics. Water and hydrocarbon adsorption capacities were determined on an IGA-3 gravimetric analyser from Hidden Isochema. Prior to the adsorption experiments the samples were outgassed at 400 °C for 12 h.

Structure refinement

The XRD patterns of the calcined ITQ-29 zeolites were measured at room temperature (22 °C) on a Philips X'Pert diffractometer with Bragg–Brentano geometry and an X'Celerator detector. Intensity data was obtained with a fixed divergence slit (0.25° for the Ge-containing sample and 0.125° for the pure-silica sample), Soller slits (incident and diffracted = 0.04° for the Ge-containing sample and 0.02° for the pure-silica sample) and Cu Kα radiation (λ = 1.5406, 1.5444 Å). Tube voltage and intensity, 45 kV and 40 mA; no secondary graphite monochromator. Step size was 0.017° = 2θ and time was 6,000 s for the Ge-containing sample and 1,450 s for the pure-silica sample.

The Rietveld refinement of Ge-ITQ-29 data was performed with program LSP7²⁵ using

a 2θ range from 13.5° to 64.2°. Number of contributing reflections is 86. No geometric restraints used. Number of structural parameters is seven. Number of profile parameters is eight, including unit-cell parameters and zero shift (−0.020° 2θ) with visually estimated background (average background = 38,000 counts). Profile function used was Pearson VII. Refined overall thermal vibration coefficient B = 6.0 Å². The residuals of the refinement were R_{wp} = 0.026, R_p = 0.018, χ² = 5.0.

The Rietveld refinement of the pure-silica ITQ-29 data with LSP7 was performed using the measured 2θ range from 5° to 75°. Number of contributing reflections is 186. No geometric restraints used. Number of structural parameters is seven. Number of profile parameters is eight, including unit-cell parameters and zero shift (−0.055° 2θ) with visually estimated background (average background = 400 counts). Profile function used was Pearson VII. Refined overall thermal vibration coefficient B = 0.7 Å². The residuals of the refinement were R_{wp} = 0.097, R_p = 0.062, R_B = 0.032, χ² = 2.8.

Catalytic tests

Cracking of 1-hexene and 4-methyl-1-pentene using the calcined Al-ITQ-29 zeolite as catalyst was performed in a fixed-bed continuous reactor at atmospheric pressure, 773 K reaction temperature, 2.9 g_{cat}/g_{C₆} s contact time and 10 s time on stream.

The methanol into olefins reaction was carried out in a fixed-bed continuous reactor at 673 K, 0.14 bar of methanol and 58 h^{−1} weight hour space velocity using the calcined Al-ITQ-29 zeolite as catalyst.

Received 7 July; accepted 27 July 2004; doi:10.1038/nature02909.

1. Barrer, R. M. *Zeolites and Clay Minerals as Sorbents and Molecular Sieves* Ch. 1, 2 and 6 (Academic, London, 1978).
2. Breck, D. W. *Zeolite Molecular Sieves: Structure, Chemistry and Use* Ch. 1, 2 (John Wiley, New York, 1974).
3. Moscou, L. in *Introduction to Zeolite Science and Practice* (eds Van Bekkum, H., Flanigen, E. M. & Jansen, J. C.) Ch. 1 (Elsevier, Amsterdam, 1991).
4. Sherman, J. D. Synthetic zeolites and other microporous oxide molecular sieves. *Proc. Natl Acad. Sci. USA* **96**, 3471–3478 (1999).
5. Barrer, R. M., Denny, P. J. & Flanigen, E. M. Molecular sieve adsorbents. US patent 3,306,922 (1967).
6. Kerr, G. T. Synthetic zeolite. US patent 3,314,752 (1967).
7. Wadlinger, R. L., Rosinski, E. J. & Plank, C. J. Synthetic zeolite. US patent 3,375,205 (1968).
8. Kerr, G. T. Chemistry of crystalline aluminosilicates II. The synthesis and properties of zeolite ZK-4. *Inorg. Chem.* **5**, 1537–1539 (1966).
9. Kuehl, G. H. Shape selective catalyst from zeolite alpha and use thereof. US patent 4,299,686 (1981).
10. Gramlich, V. & Meier, W. M. The crystal structure of hydrated NaA: A detailed refinement of a pseudosymmetric zeolite structure. *Z. Kristallogr.* **133**, 134–149 (1971).
11. Gies, H. & Marler, B. The structure-controlling role of organic templates for the synthesis of porosils in the system silica/template/water. *Zeolites* **12**, 42–49 (1992).
12. Lobo, R. F., Zones, S. I. & Davis, M. E. Structure-direction in zeolite synthesis. *J. Inclusion Phenom. Mol. Recogn. Chem.* **21**, 47–78 (1995).
13. Corma, A., Diaz-Cabañas, M. J., Rey, F., Nicolopoulos, S. & Boulahya, K. ITQ-15: The first ultralarge pore zeolite with a bi-directional pore system formed by intersecting 14- and 12-ring channels, and its catalytic implications. *Chem. Commun.* **12**, 1356–1357 (2004).
14. Corma, A., Diaz-Cabañas, M. J., Martínez-Triguero, L. J., Rey, F. & Rius, J. A large-cavity zeolite with wide pore windows and potential as an oil refining catalyst. *Nature* **418**, 514–517 (2002).
15. Corma, A., Rey, F., Valencia, S., Jordá, J. L. & Rius, J. A zeolite with interconnected 8-, 10- and 12-ring pores and its unique catalytic selectivity. *Nature Mater.* **2**, 493–497 (2003).
16. Castañeda, R., Corma, A., Fornés, V., Rey, F. & Rius, J. Synthesis of a new zeolite structure ITQ-24, with intersecting 10- and 12-membered ring pores. *J. Am. Chem. Soc.* **125**, 7820–7821 (2003).
17. Lee, H., Zones, S. I. & Davis, M. E. A combustion-free methodology for synthesizing zeolites and zeolite-like materials. *Nature* **425**, 385–388 (2003).
18. Nakano, T. & Yade, T. Synthesis, structure, and photophysical and electrochemical properties of a π-stacked polymer. *J. Am. Chem. Soc.* **125**, 15474–15484 (2003).
19. Blasco, T. et al. Preferential location of Ge in the double four-membered ring units of ITQ-7 zeolite. *J. Phys. Chem. B* **106**, 2634–2642 (2002).
20. Wang, Y., Song, J. & Gies, H. The substitution of germanium for silicon in AST-type zeolite. *Solid State Sci.* **5**, 1421–1433 (2003).
21. Olson, D. H., Cambor, M. A., Villacusa, L. A. & Kuehl, G. H. Light hydrocarbon sorption properties of pure silica Si-CHA and ITQ-3 and high silica ZSM-58. *Micropor. Mesopor. Mater.* **67**, 27–33 (2004).
22. Barrett, P. A. et al. ITQ-12: a new microporous silica polymorph potentially useful for light hydrocarbon separations. *Chem. Commun.* 2114–2115 (2003).
23. Grande, C. A., Gigola, C. & Rodríguez, A. E. Adsorption of propane and propylene in pellets and crystals of 5A zeolite. *Ind. Eng. Chem. Res.* **41**, 85–92 (2002).
24. Bouchéff, Y. et al. Formation of carbonaceous compounds from propene and isobutene over a 5A zeolite. Influence of temperature on their compositions and locations. *Ind. Eng. Chem. Res.* **36**, 3198–3204 (1997).
25. Rius, J. *LSP7: A Computer Program for Restrained Rietveld Refinements* (Institut de Ciència de Materials de Barcelona-CSIC, Catalunya, 2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Benham and K. Gallimore at Hidden Isochema for carrying out the water and hydrocarbon adsorption experiments. We also thank P. Atienzar and J. Martínez-Triguero for the photoluminescence and catalytic measurements, respectively. We are grateful to the CICYT for financial support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.C. (acorma@itq.upv.es).

Anaerobic hydrocarbon biodegradation in deep subsurface oil reservoirs

Carolyn M. Aitken¹, D. M. Jones¹ & S. R. Larter^{1,2}

¹School of Civil Engineering and Geosciences, University of Newcastle upon Tyne NE1 7RU, UK

²Department of Geology and Geophysics, University of Calgary, Calgary T2N 1N4, Canada

Biodegradation of crude oil in subsurface petroleum reservoirs is an important alteration process with major economic consequences¹. Aerobic degradation of petroleum hydrocarbons at the surface is well documented² and it has long been thought that the flow of oxygen- and nutrient-bearing meteoric waters into reservoirs was necessary for in-reservoir petroleum biodegradation³. The occurrence of biodegraded oils in reservoirs where aerobic conditions are unlikely⁴, together with the identification of several anaerobic microorganisms in oil fields⁵ and the discovery of anaerobic hydrocarbon biodegradation mechanisms^{6,7}, suggests that anaerobic degradation processes could also be responsible. The extent of anaerobic hydrocarbon degradation processes in the world's deep petroleum reservoirs, however, remains strongly contested. Moreover, no organism has yet been isolated that has been shown to degrade hydrocarbons under the conditions found in deep petroleum reservoirs⁸. Here we report the isolation of metabolites indicative of anaerobic hydrocarbon degradation from a large fraction of 77 degraded oil samples from both marine and lacustrine sources from around the world, including the volumetrically important Canadian tar sands. Our results therefore suggest that anaerobic hydrocarbon degradation is a common process in biodegraded subsurface oil reservoirs.

Aerobic biodegradation of hydrocarbons is often fast enough to be observable on a short human timescale and thus aerobic processes dominated petroleum geological thinking about subsurface petroleum biodegradation for many years. Although

aerobically degraded oilfields are described in the literature⁹, even conservative mass balances of the volumes of water needed to transport sufficient oxygen present overwhelming problems geologically in most reservoirs¹⁰. Even where meteoric water has flushed basins it does not follow that oxygen was carried to the deep reservoirs¹¹.

The first bacteria isolated from oilfield waters¹² were anaerobes and Russian workers have considered anaerobic hydrocarbon degradation in deep reservoirs to be the primary process for many years¹³. Subsurface waters are typically anaerobic and there is increasing evidence of the occurrence of viable anaerobic hydrocarbon degradation processes (for example, ref. 7 and references therein, and ref. 14), so we might reasonably expect subsurface hydrocarbon degradation to be anaerobic. However, until now, only one thermophilic anaerobic hydrocarbon degrader potentially capable of living in the deepest degraded reservoirs has been identified¹⁵ and none have been isolated which has been shown to degrade hydrocarbons under conditions found in deep petroleum reservoirs⁸. Magot and co-workers¹⁶ have shown that diverse anaerobic organisms commonly occur in oilfields. They concluded from laboratory and field studies that anaerobes were indeed responsible for most subsurface hydrocarbon degradation, and methanogenesis, an exclusively anaerobic process, seems common in oil-contaminated aquifers and reservoirs (compare with refs 8 and 17) and is a probable fate for much of the carbon dioxide produced during biodegradation⁴. These results, together with the availability of long periods of time in which degradation can occur, makes it likely that anaerobic hydrocarbon degradation is a common process in most petroleum reservoirs although direct evidence for this remains elusive¹¹.

It is only in the past twenty years that the use of hydrocarbons as substrates by anaerobic microorganisms has been investigated and the identification of metabolites and possible metabolic pathways reported (for reviews see ref. 6, and ref. 7 and references therein). The existence of indigenous thermophilic bacteria and hyperthermophilic *Archaea* in reservoirs was reported by L'Haridon and co-workers¹⁸ who also suggested that *in situ* bioconversion of crude oil fractions (probably aliphatic and aromatic hydrocarbons) was linked to sulphate reduction. A number of laboratory studies using aliphatic, aromatic, and polycyclic aromatic hydrocarbons as substrates for a variety of sulphate-reducing, de-nitrifying and

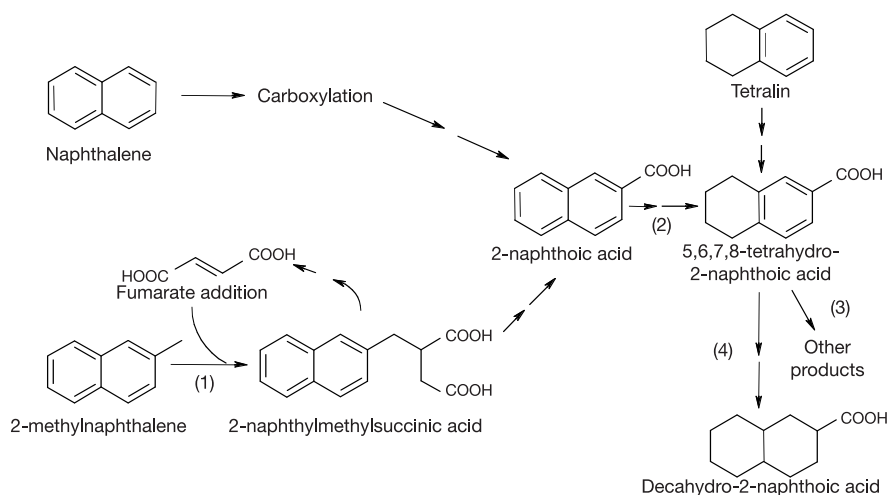


Figure 1 Part of a proposed reductive pathway for the anaerobic degradation of naphthalene, 2-methylnaphthalene and tetralin by sulphate-reducing bacteria²³. Initial carboxylation of naphthalene is followed by degradation to 2-naphthoic acid and, via a series of hydrogenation steps, to 5,6,7,8-tetrahydro-2-naphthoic acid (2), followed either by hydrogenation and further degradation steps (3) or by further hydrogenation steps (4) to give

decahydro-2-naphthoic acid (a possible dead-end metabolite). Degradation of 2-methylnaphthalene is initiated by fumarate addition to form naphthyl-2-methylsuccinic acid (1), which is further degraded to give 2-naphthoic acid. The regeneration of fumarate shown here is not shown in the original proposed pathway.

methanogenic cultures have identified succinates—formed by the addition of fumarate either to a subterminal carbon of an alkane or to an alkyl substituent of an aromatic hydrocarbon—as the initial metabolites in the degradation process^{6,7,19,20}. Succinates have also been reported as metabolites from the biodegradation of both saturated and aromatic hydrocarbons in anoxic zones of petroleum-contaminated aquifers (for example, ref. 21). 2-Naphthoic acid and reduced 2-naphthoic acids have been reported as metabolites in a laboratory study of the anaerobic degradation of naphthalene under sulphate-reducing conditions²²; in a laboratory study of the anaerobic degradation of naphthalene, 2-methylnaphthalene and tetralin, using a sulphate-reducing enrichment culture isolated from a hydrocarbon-contaminated aquifer, 2-naphthoic acid and naphthyl-2-methylsuccinic acid were identified as initial intermediates in a metabolic pathway that also identified reduced 2-naphthoic acids as further metabolites²³ (Fig. 1). Studies of anoxic zones in petroleum-contaminated aquifers have also identified 2-naphthoic acid and reduced 2-naphthoic acids as evidence of anaerobic degradation^{21,24}. We report here the identification of these metabolites in a large suite of biodegraded oils from various petroleum systems in the world.

Biodegradation in oil reservoirs requires that water, nutrients and hydrocarbons are available to sustain microbial growth. This situation is most likely to occur near the oil–water contact in the reservoir⁴. Both the transitory nature of the metabolites in the metabolic process and the extremely small amounts likely to be present contribute to increasing the difficulty of isolating and detecting such metabolites in oil samples where, with their

increased water solubility compared to hydrocarbons, they would have partitioned into the oil from material in the water-saturated portions of the reservoir. The metabolite concentrations in reservoirs containing degraded oils will also be affected by whether biodegradation is at present occurring or is inactive.

In this study the methylated acid fractions from 77 biodegraded oils, which included reservoir core solvent extracts, oil seep and tar sand samples of both marine and lacustrine shale sources from well-studied, major petroleum basins around the world were analysed by gas chromatography-mass spectrometry (GC-MS) and their spectra compared to those of known metabolite standards. Seven comparable non-biodegraded oils were similarly characterized. The oils were categorized as non-biodegraded or biodegraded and classified according to their degree of biodegradation by analysis of their hydrocarbon distributions using the criteria proposed by ref. 25. Briefly, these criteria are based on the quasi-stepwise disappearance of certain hydrocarbon classes in oils, depending on their susceptibility to biodegradation. n-alkanes are the most susceptible, followed by simply branched alkanes, whereas polycyclic alkanes are generally the most resistant. Thus, on a scale of 0 to 10, level 0 is undegraded, level 1 represents some depletion of n-alkanes, level 4 represents complete removal of n-alkanes and depletion of acyclic isoprenoids, level 8 represents destruction of the regular steranes and alteration of the hopane polycyclic hydrocarbon distributions, and so on. However, mixtures of oils at different degradation levels are often encountered in reservoirs and the scale is thus indicative rather than absolute.

Initial efforts were concentrated on the search for alkyl- and aryl-

Table 1 Metabolites indicative of anaerobic degradation found in clastic oil reservoir samples obtained from around the world

Sample	Degradation level*	Source, system, reservoir	2-naphthoic acid (p.p.m.)	5,6,7,8-tetra-hydro-2-naphthoic acid (p.p.m.)	decahydro-2-naphthoic acid (isomer 1) (p.p.m.)	decahydro-2-naphthoic acid (isomer 2) (p.p.m.)
1	1–4 (mixed)	Eocene lacustrine sourced oil, Eocene reservoir	n.d.	n.d.	0.07	0.09
2	1–4 (mixed)	Eocene lacustrine sourced oil, Eocene reservoir	n.d.	n.d.	0.18	0.16
3	1–4 (mixed)	Eocene lacustrine sourced oil, Eocene reservoir	n.d.	n.d.	n.d.	0.01
4	1–4 (mixed)	Eocene lacustrine sourced oil, Eocene reservoir	0.04	n.d.	n.d.	0.08
5	1–4 (mixed)	Eocene lacustrine sourced oil, Eocene reservoir	n.d.	n.d.	n.d.	0.11
6	1–4 (mixed)	Eocene lacustrine sourced oil, Eocene reservoir	n.d.	n.d.	n.d.	0.10
7	1–4 (mixed)	Eocene lacustrine sourced oil, Eocene reservoir	n.d.	n.d.	n.d.	0.06
8	1–4 (mixed)	Eocene lacustrine sourced oil, Eocene reservoir	n.d.	n.d.	n.d.	0.08
9	4	Cretaceous lacustrine sourced oil, Tertiary reservoir	0.33	0.30	0.06	0.28
10	4	Cretaceous lacustrine sourced oil, Tertiary reservoir	0.01	0.01	n.d.	n.d.
11	1–4 (mixed)	Miocene marine sourced oil, Miocene reservoir	0.02	n.d.	0.01	0.03
12	3	Cambrian marine sourced oil, Cretaceous reservoir	0.01	0.01	n.d.	n.d.
13	5	Eocene lacustrine sourced reservoir core, Eocene reservoir	0.30	0.05	n.d.	n.d.
14	6	Eocene lacustrine sourced reservoir core, Eocene reservoir	0.19	0.03	n.d.	n.d.
15	8	Eocene lacustrine sourced reservoir core, Eocene reservoir	0.22	0.03	n.d.	n.d.
16	2	Eocene lacustrine sourced reservoir core, Eocene reservoir	1.18	0.20	n.d.	n.d.
17	3	Eocene lacustrine sourced reservoir core, Eocene reservoir	0.67	0.06	n.d.	n.d.
18	5	Eocene lacustrine sourced reservoir core, Eocene reservoir	0.45	0.06	n.d.	n.d.
19	1–4 (mixed)	Eocene lacustrine sourced reservoir core, Eocene reservoir	1.61	0.18	n.d.	n.d.
20	1–4 (mixed)	Eocene lacustrine sourced reservoir core, Eocene reservoir	2.85	0.06	n.d.	n.d.
21	1–4 (mixed)	Eocene lacustrine sourced reservoir core, Eocene reservoir	2.08	0.11	n.d.	n.d.
22	1–4 (mixed)	Eocene lacustrine sourced reservoir core, Eocene reservoir	2.57	0.39	n.d.	0.24
23	1–4 (mixed)	Eocene lacustrine sourced reservoir core, Eocene reservoir	1.09	0.12	n.d.	n.d.
24	1–4 (mixed)	Eocene lacustrine sourced reservoir core, Eocene reservoir	9.40	0.45	n.d.	n.d.
25	1–4 (mixed)	Eocene lacustrine sourced reservoir core, Eocene reservoir	5.23	0.34	n.d.	n.d.
26	1–4 (mixed)	Eocene lacustrine sourced reservoir core, Eocene reservoir	8.15	0.31	n.d.	n.d.
27	1–4 (mixed)	Eocene lacustrine sourced reservoir core, Eocene reservoir	3.18	0.40	0.12	0.22
28	1–4 (mixed)	Eocene lacustrine sourced reservoir core, Eocene reservoir	6.70	0.46	0.13	0.27
29	1–4 (mixed)	Eocene lacustrine sourced reservoir core, Eocene reservoir	3.87	0.44	0.17	0.28
30	1–4 (mixed)	Eocene lacustrine sourced reservoir core, Eocene reservoir	8.95	0.51	0.18	0.28
31	4	Cretaceous lacustrine sourced reservoir core, Tertiary reservoir	1.92	0.23	n.d.	n.d.
32	4	Cretaceous lacustrine sourced reservoir core, Tertiary reservoir	3.04	0.31	n.d.	n.d.
33	4	Cretaceous lacustrine sourced reservoir core, Tertiary reservoir	1.07	0.21	n.d.	n.d.
34	4/5	Marine-sourced Cretaceous reservoir tar sands	0.50	n.d.	n.d.	0.01
35	7	Marine-sourced Cretaceous reservoir tar sands	0.32	nd	0.19	0.10
36	7	Marine-sourced Cretaceous reservoir tar sands	3.35	nd	0.32	0.52
37	7	Marine sourced Cretaceous reservoir tar sands	3.59	nd	0.20	0.40
38	8	Jurassic marine sourced seeps	1.75	nd	nd	0.14
39	7	Jurassic marine sourced seeps	1.46	nd	nd	0.10
40	7/8	Jurassic marine sourced seeps	1.45	0.14	nd	nd

* Assigned according to ref. 25; n.d., none detected.

Seven non-degraded oils from around the world (North Africa, Canada, China, the Middle East, the North Sea, and the Mediterranean) were also analysed and none were found to contain reduced naphthoic acids.

succinate metabolites, because these have been widely reported as anaerobic markers for both saturated and aromatic hydrocarbon degradation, both in laboratory studies and from petroleum-contaminated aquifers^{7,19–21}, and we have indeed identified such metabolites in laboratory anaerobic hydrocarbon degradation experiments²⁶. Nevertheless, attempts to confirm their presence in the oils we have studied have been inconclusive. The transient nature of these compounds, which appear to be formed in the initial step of metabolic pathways, where fumarate is regenerated as part of a cyclic process^{6,7}, suggests that they are unlikely to be detected in reservoir systems that are no longer actively degrading and, in actively degrading systems, will only be detected if sampling occurs at a currently biologically active zone. Although succinates were not reproducibly detectable, 52 of the 77 degraded oils studied were found to contain 2-naphthoic acid in concentrations ranging from 0.01 to ~10 p.p.m. and, more significantly, trace amounts of reduced 2-naphthoic acids, 5,6,7,8-tetrahydro-2-naphthoic acid and two isomers of decahydro-2-naphthoic acid were identified in 40 samples in concentrations ranging from 0.01 to ~0.5 p.p.m. (Table 1). Identification of metabolites in samples was made by comparison of mass spectral data for these samples (Fig. 2) with those of the methyl esters of standard compounds confirmed by co-injection studies with standards. To investigate the possibility that metabolites such as succinates and reduced naphthoic acids, if present, may be in the form of 'bound' biomarkers, for example, in the asphaltene fraction of degraded oils and extracts, alkaline saponification was performed on several reservoir samples before extraction and analysis of the acid fraction. Although succinates were not detected in the acid fractions of any of the saponified samples, the presence of 2-naphthoic acid and/or reduced 2-naphthoic acids was confirmed in samples in which these metabolites had previously been detected.

2-Naphthoic acid and reduced 2-naphthoic acids are formed in a reductive pathway for the anaerobic degradation of both 2-methylnaphthalene and naphthalene (Fig. 1) and as such, with abundant ubiquitous precursors in crude oils, may ultimately accumulate in sufficient quantities to be detectable in reservoir samples. The absence of reduced 2-naphthoic acids in the acid fractions of seven non-degraded oils from a variety of sources worldwide that we also analysed, and the absence of reduced 2-naphthoic acids from the acid fractions of oils that have been aerobically biodegraded in the laboratory²⁷, suggests that such compounds are indeed exclusively the products of anaerobic hydrocarbon biodegradation. Why these anaerobic metabolites were not found in all the degraded oils analysed is not known but perhaps they are only detectable in actively degrading or recently degraded reservoirs.

Although 2-naphthoic acid has been reported as both an aerobic²⁸ and an anaerobic metabolite of 2-methylnaphthalene biodegradation, 5,6,7,8-tetrahydro-2-naphthoic acid and two isomers of decahydro-2-naphthoic acid have been reported only as exclusively anaerobic metabolites^{23,24} of naphthalene and 2-methylnaphthalene biodegradation. The production of high-molecular-mass carboxylic acids by abiotic oxidation of hydrocarbons in petroleum reservoirs, such as occurs during thermochemical sulphate reduction, is in principle possible, but it only takes place at relatively high temperatures (usually $\gg 120^\circ\text{C}$)²⁹, whereas the maximum temperatures of the degraded oil reservoirs sampled in this study are all well below 85°C and high concentrations of sulphate are absent from the reservoirs. This, and the specificity of the acids seen, indicates that these are biological oxidation products.

Our results, obtained for a number of oils from a variety of reservoir sources including from the vast Canadian tar sand belt, therefore confirm that oil biodegradation in most reservoirs must

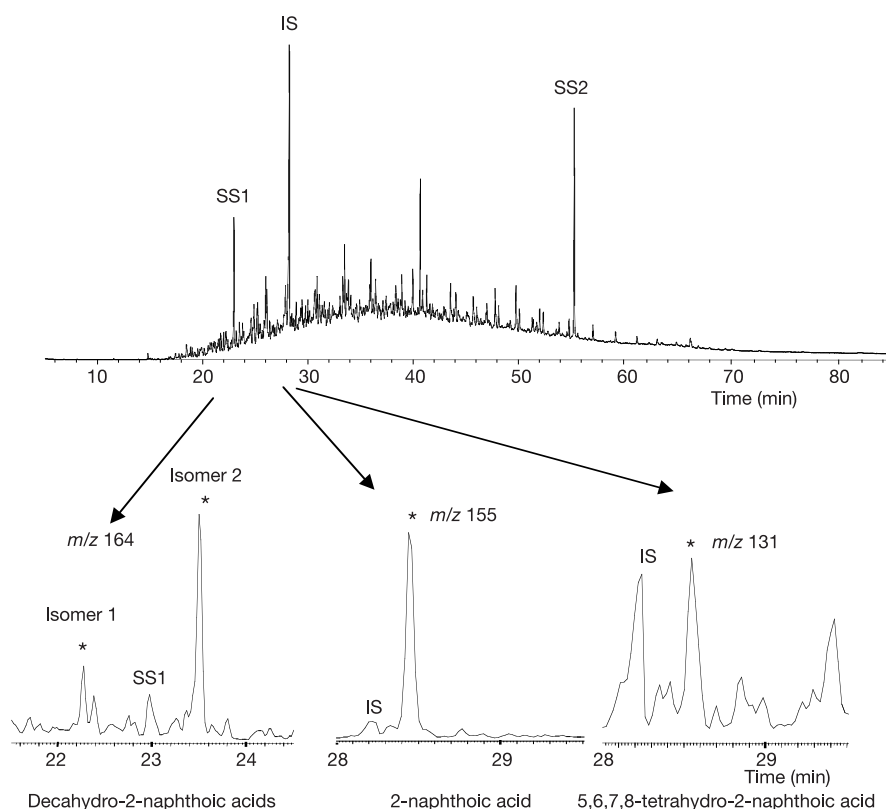


Figure 2 GC-MS data for the methylated acid fraction of a degraded oil. This shows the total ion chromatogram with recovery standard (SS) and internal standard (IS) peaks, and partial ion chromatograms for the metabolites 2-naphthoic acid (mass to charge ratio, m/z

$z = 155$); 5,6,7,8-tetrahydro-2-naphthoic acid ($m/z = 131$) and two isomers of decahydro-2-naphthoic acid ($m/z = 164$).

indeed have been anaerobic in nature at some point and indeed it seems probable that anaerobic hydrocarbon degradation is the prevailing mechanism for deep subsurface petroleum biodegradation, marking an important step in the understanding of reservoir degradation conditions. □

Methods

Total hydrocarbon fractions from crude oils were obtained by silica gel chromatography using hexane as eluent and were analysed by gas chromatography and gas chromatography-mass spectrometry (GC-MS) to provide hydrocarbon distribution patterns that provided information on the extent of biodegradation that had been suffered by the oil, as described by the scale of 0–10 proposed by ref. 25.

Acid fractions of crude oil samples were obtained using the method described in ref. 30. Briefly, ~150 mg of oil was loaded onto a strong anion exchange (SAX) quaternary amine solid phase extraction (SPE) column, and after removal of non-acid compounds, the crude carboxylic acid isolate was eluted with diethyl ether containing 2% (v/v) formic acid. This fraction was methylated with an ethereal solution of diazomethane, and cleaned up by eluting through a silica SPE column with a mixture of *n*-hexane and dichloromethane. 1-Adamantane carboxylic acid (Fluka) and 5 β cholanic acid (Sigma) were used as extraction efficiency standards and the methyl ester of 1-phenyl-1-cyclohexane carboxylic acid was used as an internal standard. Reference standards were 2-naphthoic acid (Lancaster Synthesis), 5,6,7,8-tetrahydro-2-naphthoic acid and decahydro-2-naphthoic acid. 5,6,7,8-Tetrahydro-2-naphthoic acid and decahydro-2-naphthoic acid were not commercially available and were therefore synthesized by the catalytic hydrogenation of 2-naphthoic acid using molecular hydrogen and a palladium/graphite catalyst; the identities of these compounds were confirmed by nuclear magnetic resonance spectroscopy (NMR) and by comparison of the standards' mass spectra with those in the literature.

GS-MS analysis of carboxylic acid methyl esters was performed in full scan mode on a Hewlett Packard 6890 gas chromatograph linked to a HP 5973 MSD with some confirmatory analyses carried out using a Varian 1200 GC-MS-MS system. 2-Naphthoic acid, 5,6,7,8-tetrahydro-2-naphthoic acid and decahydro-2-naphthoic acids were quantified by comparison of their peak areas in the respective *m/z* (mass-to-charge ratio) 155, 131 and 164 mass chromatograms with the peak area of the internal standard in the *m/z* = 159 mass chromatogram. Response factors of the analyte compounds were assumed to be unity, while the response factors of the surrogate standards were calculated and corrected for. Detection limits for naphthoic and reduced naphthoic acids were 0.01 p.p.m.

Received 23 February; accepted 3 August 2004; doi:10.1038/nature02922.

- Connan, J. in *Advances in Petroleum Geochemistry* (eds Brooks, J. & Welte, D. H.) Vol. 1, 299–335 (Academic, London, 1984).
- Atlas, R. M. & Bartha, R. Hydrocarbon biodegradation and oil-spill remediation. *Adv. Microb. Ecol.* **12**, 287–338 (1992).
- Palmer, S. E. in *Organic Geochemistry* (eds Macko, S. A. & Engel, M. H.) 511–534 (Plenum, New York, 1993).
- Head, I. M., Jones, D. M. & Larter, S. R. Biological activity in the deep subsurface and the origin of heavy oil. *Nature* **426**, 344–352 (2003).
- Magot, M., Ollivier, B. & Patel, B. K. C. Microbiology of petroleum reservoirs. *Antonie van Leeuwenhoek Int. J. Gen. Mol. Microbiol.* **77**, 103–116 (2000).
- Boll, M., Fuchs, G. & Heider, J. Anaerobic oxidation of aromatic compounds and hydrocarbons. *Curr. Opin. Chem. Biol.* **6**, 604–611 (2002).
- Widdel, F. & Rabus, R. Anaerobic biodegradation of saturated and aromatic hydrocarbons. *Curr. Opin. Biotechnol.* **12**, 259–276 (2001).
- Röling, W. F. M., Head, I. M. & Larter, S. R. The microbiology of hydrocarbon degradation in subsurface petroleum reservoirs: perspectives and prospects. *Res. Microbiol.* **154**, 321–328 (2003).
- Evans, C. R., Rogers, M. A. & Bailey, N. J. L. Evolution and alteration of petroleum in Western Canada. *Chem. Geol.* **8**, 147–170 (1971).
- Hörstad, I., Larter, S. R. & Mills, N. A quantitative model of biological petroleum degradation within the Brent Group reservoir in the Gullfaks field, Norwegian North Sea. *Org. Geochem.* **19**, 107–117 (1992).
- Larter, S. R. *et al.* The controls on the composition of biodegraded oils in the deep subsurface. Part 1: biodegradation rates in petroleum reservoirs. *Org. Geochem.* **34**, 601–613 (2003).
- Bastin, E. S. The presence of sulphate reducing bacteria in oil field waters. *Science* **63**, 21–24 (1926).
- Kartsev, A. A., Tabasarsanskii, Z. A., Subbota, M. I. & Mogilevski, G. A. *Geochemical Methods of Prospecting and Exploration for Petroleum and Natural Gas* (English translation edited by Witherspoon, P. A. & Romey, W. D.) (Univ. California Press, Berkeley, California, 1959).
- Zengler, K., Richnow, H. H., Rosselló-Mora, R., Michaelis, W. & Widdel, F. Methane formation from long-chain alkanes by anaerobic microorganisms. *Nature* **401**, 266–269 (1999).
- Rueter, P. *et al.* Anaerobic oxidation of hydrocarbons in crude oil by new types of sulfate-reducing bacteria. *Nature* **372**, 455–458 (1994).
- Magot, M., Connan, J. & Crolet, J.-L. Les bactéries des gisements pétroliers. *La Recherche* **25** (268), 936–937 (1994).
- Bennett, P. C., Siegel, D. E., Baedeker, M. J. & Hult, M. F. Crude oil in a shallow sand and gravel aquifer. I. Hydrogeology and inorganic geochemistry. *Appl. Geochem.* **8**, 529–549 (1993).
- L'Haridon, S., Reysenbach, A.-L., Glénat, P., Prieur, D. & Jeanthon, C. Hot subtterranean biosphere in a continental oil reservoir. *Nature* **377**, 223–224 (1995).
- Elshahed, M. J., Gieg, L. M., McNerney, M. J. & Sufliata, J. M. Signature metabolites attesting to the *in situ* attenuation of alkylbenzenes in anaerobic environments. *Environ. Sci. Technol.* **35**, 682–689 (2001).
- Rios-Hernandez, L. A., Gieg, L. M. & Sufliata, J. M. Biodegradation of an alicyclic hydrocarbon by a sulfate-reducing enrichment from a gas condensate-contaminated aquifer. *Appl. Environ. Microbiol.* **69**, 434–443 (2003).
- Gieg, L. M. & Sufliata, J. M. Detection of anaerobic metabolites of saturated and aromatic

- hydrocarbons in petroleum-contaminated aquifers. *Environ. Sci. Technol.* **36**, 3755–3762 (2002).
- Zhang, X., Sullivan, E. R. & Young, L. Y. Evidence for aromatic ring reduction in the biodegradation pathway of carboxylated naphthalene by a sulfate reducing consortium. *Biodegradation* **11**, 117–124 (2000).
- Annweiler, E., Michaelis, W. & Meckenstock, R. U. Identical ring cleavage products during anaerobic degradation of naphthalene, 2-methylnaphthalene, and tetralin indicate a new metabolic pathway. *Appl. Environ. Microbiol.* **68**, 852–858 (2002).
- Phelps, C. D., Battistelli, J. & Young, L. Y. Metabolic biomarkers for monitoring anaerobic naphthalene biodegradation *in situ*. *Environ. Microbiol.* **4**, 532–537 (2002).
- Peters, K. E. & Moldowan, J. M. *The Biomarker Guide* (Prentice Hall, New York, 1993).
- Aitken, C. M. *Identification of Non-Hydrocarbon Metabolites of Deep Subsurface Anaerobic Petroleum Hydrocarbon Biodegradation* PhD thesis, Univ. Newcastle-upon-Tyne, UK (2004).
- Watson, J. S. *Hydrocarbon and Carboxylic Acid Compositions of Crude Oil Biodegraded in Marine Systems* PhD thesis, Univ. Newcastle-upon-Tyne, UK (1999).
- Van der Linden, A. C. & Thijssse, G. J. E. The mechanisms of microbial oxidations of petroleum hydrocarbons. *Adv. Enzymol.* **27**, 469–546 (1965).
- Seewald, J. S. Organic-inorganic interactions in petroleum-producing sedimentary basins. *Nature* **426**, 327–333 (2003).
- Jones, D. M., Watson, J. S., Meredith, W., Chen, M. & Bennett, B. Determination of naphthenic acids in crude oils using non-aqueous ion exchange solid-phase extraction. *Anal. Chem.* **73**, 703–707 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank I. Head, B. Bennett and H. Huang for comments and supporting data, and also the members of the BACCHUS biodegradation consortium for support, discussions and permission to publish. BACCHUS members include Norsk Hydro, Shell, Enterprise, Petrobras, TotalFinaElf, Exxon Mobil, JNOC, ConocoPhillips, BP Amoco and ChevronTexaco. We also thank H. Wilkes for discussions and the gift of succinate standards. We acknowledge support from the Natural Environment Research Council for research funds, including JREI awards and S.R.L. acknowledges additional support from NSERC and AIF.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.M.J. (martin.jones@ncl.ac.uk).

Complex organic chemical balms of Pharaonic animal mummies

Stephen A. Buckley*, Katherine A. Clark & Richard P. Evershed

Organic Geochemistry Unit, Bristol Biogeochemistry Research Centre, School of Chemistry, Cantock's Close, Bristol BS8 1TS, UK

* Present address: Department of Archaeology, University of York, King's Manor, York YO1 7EP, UK

Millions of votive mummies of mammals, birds and reptiles were produced throughout ancient Egypt, with their popularity increasing during the reign of Amenhotep III (1400 BC) and thereafter. The scale of production has been taken to indicate that relatively little care and expense was involved in their preparation compared with human mummies^{1–3}. The accepted view is that animals were merely wrapped in coarse linen bandages and/or dipped in 'resin' before death^{2–4}. However, as with human mummification there was a range of qualities of treatments, and visual inspection of animal mummies suggests that the procedures used were often as complex as those used in humans (for example, evisceration and elaborate bandaging). Moreover, the ancient Egyptians treated animals with great respect, regarding them both as domestic pets and representatives of the gods; for example, the cat symbolized the goddess Bastet; the hawk, Horus; the ibis, Thoth, and so on. We report here the results of chemical investigations of tissues and wrappings from Pharaonic cat, hawk and ibis mummies using gas chromatography, gas chromatography-mass spectrometry, thermal desorption-gas chromatography-mass spectrometry and pyrolysis-gas chromatography-mass spectrometry^{5,6}. The analyses reveal the presence of highly complex mixtures of

n-alkyl and cyclic biomarker components characteristic of fats, oils, beeswax, sugar gum, petroleum bitumen, and coniferous, *Pistacia* and possibly cedar resins. The mixture of balms is of comparable complexity to those used to mummify humans from the same period^{6–8}.

One essential aspect of the mummification process was the treatment of the corpse and associated viscera with various organic embalming agents. Previous reports have shown that a variety of natural products possessing a range of preservative properties were applied during mummification, including beeswax^{6–8}, animal fats and plant oils^{6–8}, plant resins^{6–9}, petroleum bitumen^{7–9} and essential oils^{10,11}. Thus, a further criterion on which to assess the degree of care and sophistication employed in the mummification of animal mummies would be to determine the chemical composition of their balms. The specimens investigated herein (Fig. 1) are representatives of the latest mummy-making era of Pharaonic Egypt (XXIII dynasty (818–715 BC) to the XXX dynasty (380–343 BC)). As in our previous report on human mummies⁷, we focused on diagnostic marker compounds that are resistant to degradation and can be related to specific natural products comprising the original embalming materials.

The embalming materials identified in the animal mummies (Table 1) are either sugar gum (that is, polysaccharide-based plant exudate, such as gum arabic) and/or lipid based. The fatty acid compositions of their extracts are indicative of both animal (low

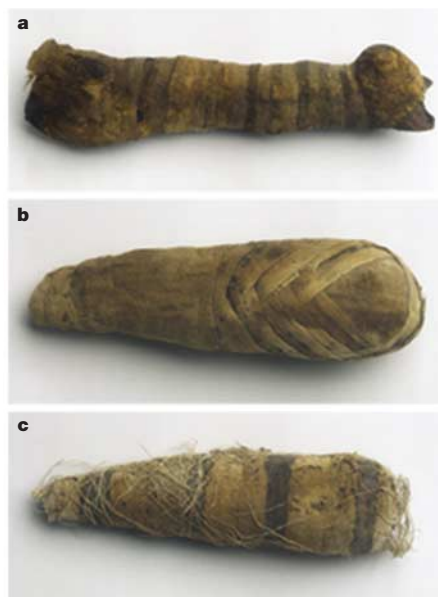


Figure 1 Photographs of animal mummies used in this investigation. **a**, Cat. **b**, Hawk. **c**, Ibis. Images were obtained after taking milligram to sub-milligram samples of balms and wrappings indicated in Table 1, showing that sampling for micro-scale analyses was essentially non-destructive.

Table 1 Provenance and date of mummies, origin of balms and their chemical composition

Mummy (museum number)	Date (yr BC) (dynasties)	Provenance	Sample description†	Inferred components of embalming 'resin'‡	Relative abundance (%)§
Hawk* (52.55.46)	818–664 (XXIII–XXV)	Tarkhan	'Resin' on wrapping underneath jaw/mandible	Fat/oil ^a Wax (beeswax?) ^{i,n}	90 10
			'Resin' on back, base of neck (head of spine)	Fat/oil ^b Wax (beeswax?)	90 10
			'Resin' on wrapping covering right breast	Fat/oil ^c Wax (beeswax?)	90 10
Hawk* (52.55.47)	818–664 (XXIII–XXV)	Tarkhan	'Resin'-soaked linen above right eye	Fat/oil ^d Wax ^{k,o}	30 70
Cat* (56.22.224)	664–343 (XXVI–XXX)	Beni Hassan	Blackened wrapping from base of mummy	Animal fat ^e	60
				Cedar resin (?)	4
				Pistacia resin	4
				Balsam/umbelliferae	1
				Beeswax ^{k,p}	30
				Gum resin (myrrh?)	1
			Detached 'resin'-soaked wrapping (location 1)	Animal fat ^f	61
				Cedar resin (?)	5
				Balsam/umbelliferae	1.5
				Beeswax ^{l,q}	31
				Gum resin (myrrh?)	1.5
				Petroleum bitumen	Trace
			Detached 'resin'-soaked wrapping (location 2)	Animal fat ^g	64
				Cedar resin (?)	2
				Pistacia resin	0.2
				Balsam/umbelliferae	0.5
				Beeswax ^{m,r}	33
			Red material in right ear	Gum resin (myrrh?)	0.5
				A sugar gum	95
				Plant oil ^h	5
				Beeswax	0.5
Ibis* (1969.112.42)	664–343 (XXVI–XXX)	Sakkara	'Resin'-soaked wrapping, covering right breast	A sugar gum Plant oil Wax	100 Trace Trace
			'Resin'-soaked wrapping, covering left breast	A sugar gum Plant oil Wax	100 Trace Trace

*Liverpool Museum.

†The term 'resin' denotes physical appearance and does not presuppose any chemical composition or biological origin.

‡Superscript letters (a–r) refer to the histograms in Fig. 2.

§Percentage of relative abundance on the basis of absolute concentrations in lipid extract, determined by gas chromatography on the basis of internal standards added at the extraction stage (thermal desorption was taken into account where appropriate). Compositions do not imply that they were the original formulations of the embalmers, owing to possible chemical changes over time.

||Diagnostic steranes and hopanes present in trace concentrations. Concentration of bitumen not determined owing to chemical complexity.

abundance of $C_{16:0}$ compared with $C_{18:0}$) and plant (high abundance of $C_{16:0}$ compared with $C_{18:0}$) acyl lipid origins (see Fig. 2a–h). The ibis mummy treatment contains only trace amounts of fatty acids, being largely sugar based, although the high $C_{16:0}$ to $C_{18:0}$ ratios for the two samples seem to indicate a plant origin. In contrast, the three samples of wrappings taken from the mummified cat have $C_{16:0}$ to $C_{18:0}$ ratios of ~ 1.5 , which taken together with the presence of cholesterol derivatives and triacylglycerols (with a high abundance of the $C_{18:0}$ fatty acyl moiety; Fig. 2) confirm an animal origin; a contribution from endogenous body fats cannot be ruled out. Red material that had been packed into the ears of the cat has a quite different $C_{16:0}$ to $C_{18:0}$ ratio (3.7) indicating a possible plant origin. The ratios of $C_{16:0}$ and $C_{18:0}$ in all the samples from the hawk mummies are very similar (~ 1.5), which suggests an animal fat origin. These samples all contained an unusually high abundance of 9,10-dihydroxyoctadecanoic acid, with the differing abundances of the *threo* and *erythro* isomers possibly reflecting differing sources or extents of stereomutation during the oxidation of the precursor oleic acid¹². As in the previous study of human mummies, the results suggest that animal fats and plant oils were used as a less costly base on which to apply more exotic ingredients⁷.

The presence of monosaccharides in the solvent extracts, and abundant sugar markers (levoglucosan, furan and pyran derivatives) in the thermal desorption–gas chromatography–mass spectrometry

analyses, indicate the presence of a hydrolyzed plant sugar gum in the ibis mummy. Because both furanose and pyranose sugars were present, their origin cannot be the cellulose-based wrappings. Unfortunately, the low abundance of sugars in the extracts makes their precise origin difficult to assign; however, the identification of a sugar gum is consistent with a previous report of the use of a “gum” to secure mummy wrappings¹³. Analogous sugar gum markers were also detected in the red packing and wrappings of the cat mummy.

Conifer resin was identified in the three impregnated wrappings from the mummified cat. The normally abundant diterpenoids, particularly dehydroabietic acid, are highly oxidized, such that 7-oxodehydroabietic acid is most abundant with lesser amounts of 15-hydroxy-7-oxodehydroabietic acid (Fig. 3). Importantly, these diterpenoids are accompanied by sesquiterpenoids indicating that essential oil(s), such as cedar oil, was a component of the balm. Although tetramethylhexahydrobenzocycloheptadiene is characteristic of cedar resin and has been identified in a number of Egyptian vessels¹⁰, the presence of this compound could not be unambiguously confirmed. However, the mass spectra of later eluting components of the neutral fraction exhibit $M^+ \cdot 216$ and 218, consistent with oxidation products of the aforementioned hydrocarbon. The precursor benzocycloheptadiene would be susceptible to autoxidation, and given the presence of the predominantly

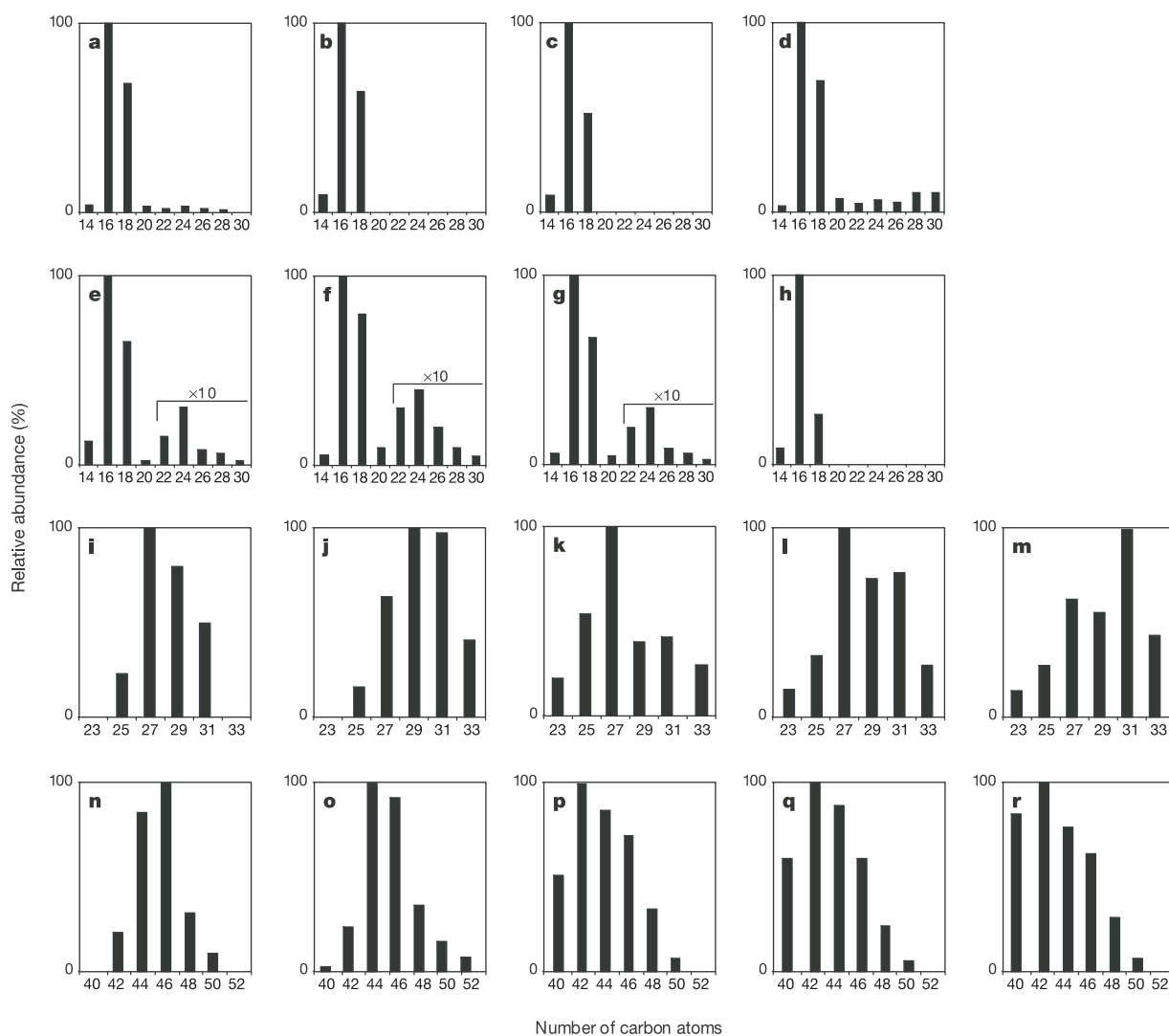


Figure 2 Histograms showing the distributions of alkyl lipids in balms of animal mummies. Fatty acids characteristic of plant oils and animal fats (a–h), and *n*-alkanes (i–m) and wax esters (n–r) present in balms containing beeswax as a significant component.

oxidized diterpenoids in the extract, then the oxidized forms of sesquiterpenoids would be expected to dominate.

The subject of mummification was written about extensively in the early part of the twentieth century^{1,14}, with the opinion that 'cedar oil' was in fact not true cedar, but a juniper. Unfortunately, our results do not confirm or refute this assertion, but do emphasize the problems of oxidation during either the preparation of balms or the deterioration of these materials over time, hindering confident assignments of this intriguing component. Hence, the report by classical author Diodorus Siculus on feline embalming¹⁵, "... it is embalmed with cedria and the other substances which have the virtue of preserving bodies..." remains tantalizingly to be unambiguously confirmed.

However, triterpenoids were observed in two of the wrappings from the cat mummy (Table 1 and Fig. 3) including moronic acid as the dominant component, with lesser amounts of dammaranes, along with nor- α -amyrone (3-oxo-28-nor-urs-12-ene) and nor- β -

amyrone (3-oxo-28-nor-olean-12-ene). The wrappings do not contain the euphane markers, isomasticadienonic acid and masticadienonic acid, identified in previous studies^{7,8,16}, but the presence of the diagnostic and relatively stable moronic acid clearly indicates that *Pistacia resin*¹⁷ was used in embalming the cat.

The use of beeswax, characterized chemically by *n*-alkanes (C₂₅ to C₃₃; Fig. 2i–m), wax esters (C₄₀ to C₅₀; Fig. 2n–r) and hydroxy wax esters (C₄₂ to C₅₄), is conclusively shown in the cat mummy. The solvent soluble extracts comprise 30–33% w/w beeswax (Fig. 4). The fourth sample from the wrappings of the cat mummy contains only traces of C₂₅ to C₃₁ *n*-alkanes, with no detectable wax esters, although the predominance of the C₂₇ alkane indicates a probable beeswax origin. The sample taken from the hawk (museum number 52.55.46) also contains a significant amount of wax (~10% w/w), the wax ester distribution being similar to that of beeswax but with the C₄₀ homologue being greatly reduced in abundance. The presence of *n*-alkanes (C₂₅ to C₃₁ with an odd-over-even predominance and

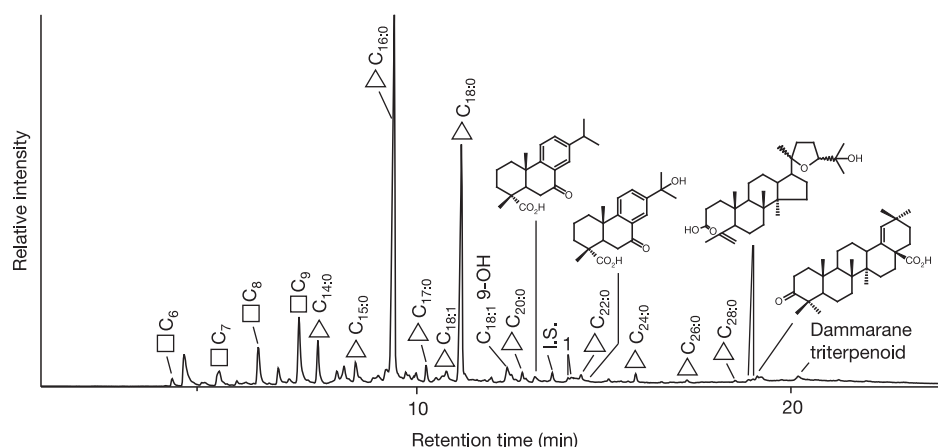


Figure 3 Reconstructed gas chromatography mass spectrometry (GC-MS) total ion chromatogram (TIC) of the trimethylsilylated acid fraction of blackened wrapping from the base of the mummy of the Late Pharaonic period cat. Peak identities ('x' indicates carbon chain length): open triangles, C_x:0 indicates saturated fatty acids; open squares, C_x indicates α,ω -dicarboxylic acids. C_{18:1} 9-OH is 9-hydroxy-10-octadecenoic acid and

1 = C_{18:0} 9,10-di-OH (*t* + *e*), which is 9,10-dihydroxyoctadecanoic acid (*threo* and *erythro* isomers). Also shown are the structures of two diterpenoid, and two triterpenoid, acids identified: 7-oxodehydroabietic acid; 15-hydroxy-7-oxodehydroabietic acid; 20,24-epoxy-25-hydroxy-3,4-seco-4(28)-dammaren-3-oic acid (two isomers); and moronic acid. I.S. is the internal standard (*n*-heneicosanoic acid).

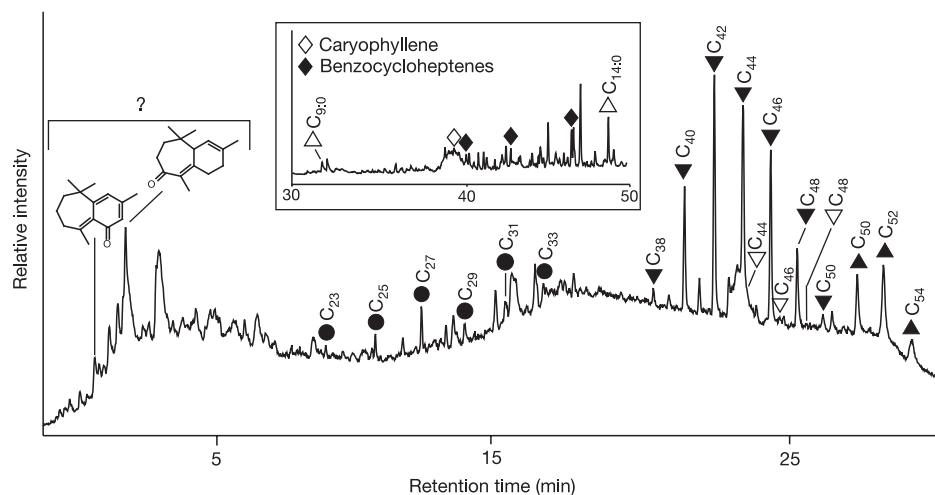


Figure 4 Reconstructed GC-MS TIC of the trimethylsilylated neutral fraction of blackened wrapping from the base of the mummy of the Late Pharaonic period cat. Inset shows thermal desorption-GC-MS (TD-GC-MS) of the same sample. Peak identities ('x' indicates carbon chain length): filled circles, C_x indicates *n*-alkanes; filled inverted triangles, C_x indicates wax esters; open inverted triangles, C_x indicates hydroxy wax

esters; filled triangles, C_x indicates triacylglycerols; I.S. is the internal standard (*n*-tetratriacontane). Also shown are the structures of two oxidized benzocycloheptenes tentatively identified. The inset displays a reconstructed TIC of the TD-GC-MS of this sample, showing the *Pistacia* sesquiterpenoid, caryophyllene, and the benzocycloheptenes tentatively identified.

maximizing at C_{27}), the dominance of the $C_{16:0}$ acyl group in the wax esters and the saturated long-chain fatty acids in the $C_{22:0}$ to $C_{28:0}$ carbon number range, with $C_{24:0}$ predominating, all point to the presence of beeswax, albeit in a degraded state^{7,8,18–20}. A similarly degraded beeswax was also seen in the other mummified hawk. The two samples from the ibis mummy also contain alkanes and wax

esters indicative of a wax, but their low concentrations preclude a firm conclusion regarding its origin.

The gas chromatogram of the neutral fraction of the ‘resin’-coated bandage taken from the cat mummy displayed a notable ‘hump’ (Fig. 4), which is indicative of a thermally mature petroleum product (that is, ‘true’ bitumen). The hydrocarbon fraction⁶ contained characteristic bitumen-derived triterpanes and steranes, although their concentrations were about three orders of magnitude lower than those of other lipids. The source of this bitumen, determined using the ratios of the most characteristic bitumen biomarkers, suggests that it most closely resembles that from Abu Durba, Gulf of Suez, although a mixture with Dead Sea bitumen or another source is possible (Fig. 5)²¹. Gebel Zeit, another source thought to have been available to the Egyptians, has ratios markedly different to those calculated for the cat and can therefore be discounted as the source.

The identification of a range of natural sesqui-, di- and triterpenoids, wax esters, fatty acids and bitumen biomarkers demonstrate that commodities considered as exotic were used in animal mummification¹. The complexity of the mixtures is analogous to those previously detected in human mummies^{6–9,22–24}. The choice of natural products employed suggests that the ancient Egyptians had an appreciation of their preservative properties⁷. These findings provide further evidence for animal mummies being prepared using procedures as sophisticated as those employed in human mummification. □

Received 13 May; accepted 13 July 2004; doi:10.1038/nature02849.

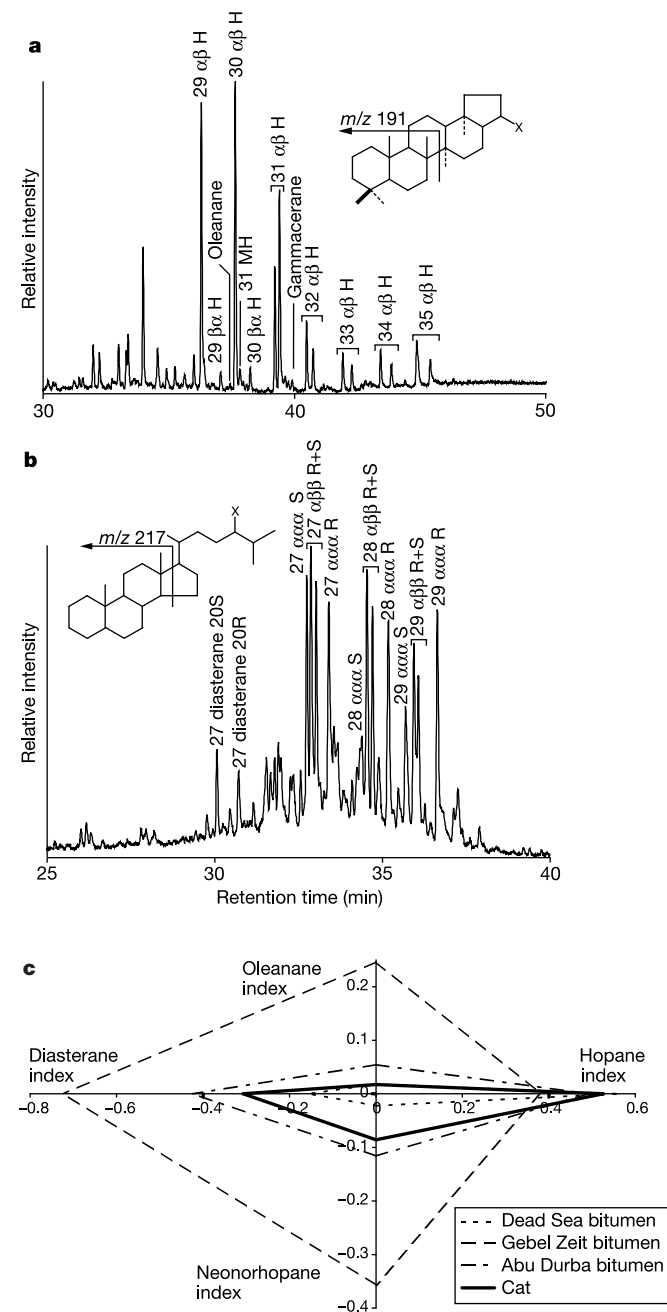


Figure 5 GC-MS selected ion monitoring mass chromatograms and star diagrams. **a, b**, Mass-to-charge ratio (m/z) 191 (**a**) and 217 (**b**) GC-MS selected ion monitoring mass chromatograms of saturated hydrocarbon fraction of ‘resin’-coated bandages taken from the cat mummy. **c**, Star diagrams for reference petroleum bitumen and cat extract. Indices are calculated on the basis of peak heights as decimal fractions: oleanane index = oleanane/(oleanane + $17\alpha,21\beta$ (H)-hopane); hopane index = $17\alpha,21\beta$ (H)-29-pentakishomohopanes/($17\alpha,21\beta$ (H)-29-pentakishomohopanes + $17\alpha,21\beta$ (H)-29-trishomohopanes); neonorhopane index = 18α (H)-30-neonorhopane/(18α (H)-30-neonorhopane + 17α (H), 21β (H)-30-neonorhopane); diasterane index = $13\beta,17\alpha$ -diacholestane 20S/($5\alpha,14\beta,17\beta$ -cholestane 20S + $13\beta,17\alpha$ -diacholestane 20S). Reference values taken from ref. 21.

- Lucas, A. *Ancient Egyptian Materials and Industries*, 4th edn (Histories and Mysteries of Man, London, 1989).
- Malek, J. *The Cat in Ancient Egypt* (British Museum, London, 1997).
- Brier, B. *Egyptian Mummies* (Michael O’ Mara Books, London, 1996).
- Ikram, S. & Dodson, A. *The Mummy in Ancient Egypt* (Thames and Hudson, London, 1998).
- Buckley, S. A., Stott, A. W. & Evershed, R. P. Studies of organic residues from ancient Egyptian mummies using high temperature gas chromatography mass spectrometry and sequential thermal desorption gas chromatography mass spectrometry and pyrolysis gas chromatography mass spectrometry. *Analyst* **124**, 443–452 (1999).
- Connan, J. & Dessort, D. Du bitumen dans les baumes des momies égyptiennes (1295 av JC–300 ap JC): détermination de son origine et évaluation de sa quantité. *C.R. Acad. Sci. Ser. II* **312**, 1445–1452 (1991).
- Buckley, S. A. & Evershed, R. P. Organic chemistry of embalming agents in Pharaonic and Graeco-Roman mummies. *Nature* **413**, 837–841 (2001).
- Colombini, M. P., Modugno, F., Silvano, F. & Onor, M. Characterization of the balm of an Egyptian mummy from the 7th century B.C. *Stud. Conserv.* **45**, 19–29 (2000).
- Maurer, J., Möhring, T., Rullkötter, J. & Nissenbaum, A. Plant lipids and fossil hydrocarbons in embalming material of Roman period mummies from the Dakhleh Oasis, Western Desert, Egypt. *J. Archaeol. Sci.* **29**, 751–762 (2002).
- Serpico, M. & White, R. in *Chemical Analysis of Coniferous Resins from Ancient Egypt using Gas Chromatography/Mass Spectrometry (GC/MS)* (ed. Eyre, C.) 1037–1048. (Proc. Seventh Intern. Congr. Egyptol., Leuven, 1998).
- Koller, J., Baumer, U., Kaup, Y., Schmid, M. & Weser, U. Ancient materials - Analysis of a Pharaonic embalming tar. *Nature* **425**, 784 (2003).
- Gunstone, F. D., Harwood, J. L. & Padley, F. B. *The Lipid Handbook* (Chapman and Hall, London, 1986).
- Herodotus, *The Histories* (trans. De Sélincourt, A.) (Penguin, London, 1996).
- Lucas, A. “Cedar”-tree products employed in mummification. *J. Egypt. Archaeol.* **XVII**, 13–21 (1931).
- Diodorus, *The Library of History* (trans. Oldfather, C. H.) (Heinemann, London, 1935).
- Mills, J. S. & White, R. The identity of the resins from the Late Bronze-Age shipwreck at Ulu-Burun (Kas). *Archaeometry* **31**, 37–44 (1989).
- Serpico, M. & White, R. in *Ancient Egyptian Materials and Technology* (eds Nicholson, P. T. & Shaw, I.) 430–474 (Cambridge Univ. Press, Cambridge, 2000).
- Evershed, R. P., Vaughan, S. J., Dudd, S. N. & Soles, J. S. Fuel for thought? Beeswax in lamps and conical cups from late Minoan Crete. *Antiquity* **71**, 979–985 (1997).
- Evershed, R. P., Dudd, S. N., Anderson-Stojanovic, V. R. & Gebhard, E. R. New chemical evidence for the use of combed ware pottery vessels as beehives in ancient Greece. *J. Archaeol. Sci.* **30**, 1–12 (2003).
- Regert, M., Colinart, S., Degrand, L. & Decavallas, O. Chemical alteration and use of beeswax through time: Accelerated ageing tests and analysis of archaeological samples from various environmental contexts. *Archaeometry* **43**, 549–569 (2001).
- Harrell, J. A. & Lewan, M. D. Sources of mummy bitumen in ancient Egypt and Palestine. *Archaeometry* **44**, 285–293 (2002).
- Connan, J. Use and trade of bitumen in antiquity and prehistory: molecular archaeology reveals secrets of past civilizations. *Phil. Trans. R. Soc. Lond. B* **354**, 33–50 (1999).
- Rullkötter, J. & Nissenbaum, A. Dead-Sea asphalt in Egyptian mummies - molecular evidence. *Naturwissenschaften* **75**, 618–621 (1988).
- Koller, J., Baumer, U., Kaup, Y., Etschpüler, H. & Weser, U. Embalming was used in Old Kingdom. *Nature* **391**, 343–344 (1998).

Acknowledgements We thank J. Hayward of Liverpool Museum for making the samples available to us. Thanks also go to J. Carter, I. Bull, A. Gledhill and B. van Dongen for their technical assistance. The National Gallery Scientific Laboratory is thanked for the gift of cedar oil. The NERC provided financial support for mass spectrometry facilities and a studentship to K.A.C.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.P.E. (r.p.evershed@bristol.ac.uk).

Early brain growth in *Homo erectus* and implications for cognitive ability

H. Coqueugniot¹, J.-J. Hublin², F. Veillon³, F. Houët¹ & T. Jacob⁴

¹UMR 5199-PACEA, Laboratoire d'Anthropologie des Populations du Passé, Université Bordeaux 1, avenue des Facultés, 33405 Talence cedex, France

²Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04130 Leipzig, Germany

³Service de Radiologie I, Hôpital Hautepierre, avenue Molière, 67200 Strasbourg, France

⁴Department of Physical Anthropology, Gadjah Mada University, College of Medicine, Yogyakarta, Indonesia

Humans differ from other primates in their significantly lengthened growth period. The persistence of a fetal pattern of brain growth after birth is another important feature of human development¹. Here we present the results of an analysis of the 1.8-million-year-old Mojokerto child (Perning 1, Java), the only well preserved skull of a *Homo erectus* infant, by computed tomography. Comparison with a large series of extant humans and chimpanzees indicates that this individual was about 1 yr

(0–1.5 yr) old at death and had an endocranial capacity at 72–84% of an average adult *H. erectus*. This pattern of relative brain growth resembles that of living apes, but differs from that seen in extant humans. It implies that major differences in the development of cognitive capabilities existed between *H. erectus* and anatomically modern humans.

Dental microstructure has been recently used² to determine when in the course of hominid evolution a modern human pattern of dental maturation appeared. Representatives of *H. erectus* have been shown to display a shorter period of dental development, suggesting that a modern human growth pattern evolved more recently. Another important aspect of human growth is 'secondary altriciality'. In most primates, brain growth slows down rapidly after birth¹ whereas hominids have to solve the evolutionary challenge of developing a large brain under substantial physiological, obstetrical and locomotor constraints³. An adaptive solution has been reached by giving birth to offspring with relatively small brains compared with adult brain size. Whereas *Macaca* newborns display an endocranial volume equivalent to 70% of adult size¹, the modern human brain represents only 25% of its adult size at birth and continues to grow at its fast fetal rate during the first year of life. At 1 yr of age the human brain is 50% of its adult size and at 10 yr 95% of the adult brain size is achieved. At birth, apes display an intermediate condition, with an endocranial volume approximately 40% of adult size in the common chimpanzee⁴, with 80% of the adult volume being reached by the end of the first year.

Secondary altriciality has social consequences: modern human children require many years of parental support. It also influences the development of cognitive abilities. Most of human brain growth takes place in an 'enriched environment', while the individual is already interacting with the extra-maternal environment^{5–7}. A prolonged interaction between peripheral somatic areas and developing related sensori-motor cortical areas could be one condition for the development of spoken language. When this important adaptation of the genus *Homo* appeared during the course of

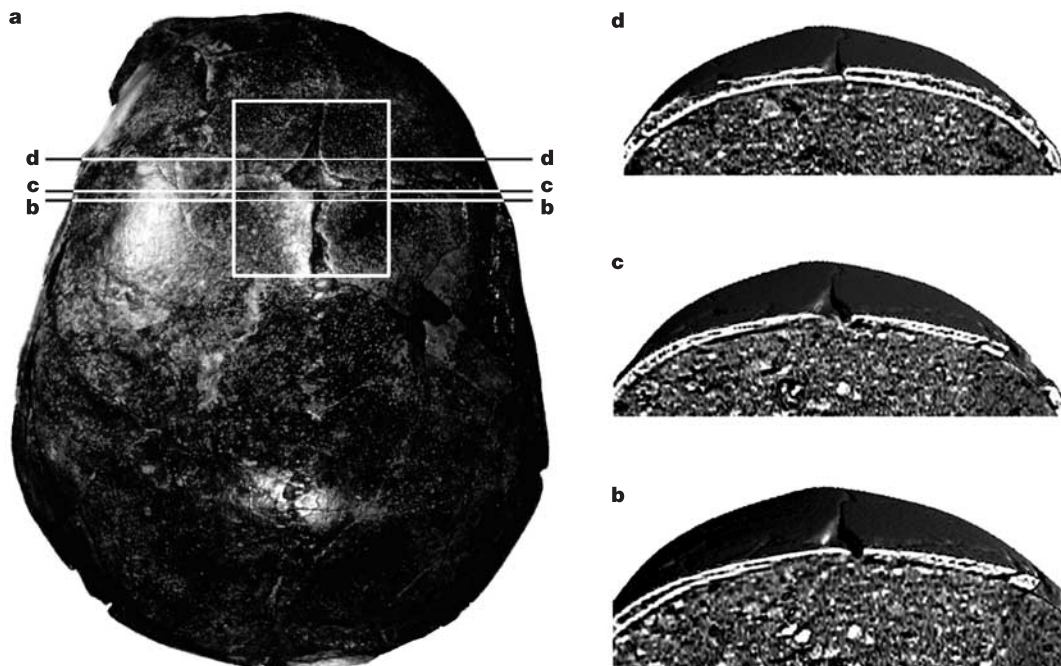


Figure 1 Superior view of the Mojokerto specimen (a) and three-dimensional reconstructions from axial CT scans of the anterior part of the skull (b–d). The location of the coronal cuts are indicated on a. The coronal cuts in b and c, located immediately behind the bregma, display the gap between the two parietals reduced to one compact

table on both sides. The large depression located anteriorly to the bregma on the frontal visible on d results from damage of the outer table. Three-dimensional reconstructions of the skull were made with Voxel-man software (University of Hamburg)²⁸. The photograph in a is from the Senckenberg Museum.

evolution has been much debated^{1–3,8–10}. So far the fossil record has not provided clear evidence of when secondary altriciality developed. In this paper we address this question by analysing the only well preserved brain case of a *H. erectus* infant.

The Mojokerto child was discovered in the Pucangan layers (Perning, Java, Indonesia) in 1936. A hornblende sample extracted from the pumice-bearing layer where the specimen was purportedly found was dated to 1.81 ± 0.04 million years (Myr) ago by the $^{40}\text{Ar}/^{39}\text{Ar}$ method¹¹. Although this age has been contested, Huffman¹² convincingly argued, on the basis of archival research and primary fieldwork, that the Perning child was found *in situ* in the upper

Pucangan Formation and that the samples used by Swisher *et al.*¹¹ are appropriate as age estimates of this fossil. Consequently, this child may well represent the earliest evidence of hominids in Indonesia, a hypothesis reinforced by the discovery of other hominids of similar age outside of Africa^{13,14}.

The Mojokerto specimen is represented by an almost complete calvaria. Because the specimen lacks dental remains, the determination of its individual age is based on the relative state of development of various cranial structures, and has varied among authors from 18 months to 8 yr old based on modern human standards^{15–18}. The latest estimate of the age of the specimen¹⁹ provided an evaluation of the developmental age using modern human standards, and arrived at an age of between 4 and 6 yr. This evaluation was largely based on the assumption that the anterior fontanelle was closed and on the external aspect of the temporal bone.

By computed tomography (CT) scanning of the Mojokerto calvaria we have been able to re-assess its developmental age. We analysed the maturation of three anatomical areas: the tympanic plate, the bregmatic area of the cranial vault and the subarcuate fossa (a possible criterion of maturation, the sutura mendosa, was not preserved). The development of these three features was scored by known chronological age and by dental stage in a modern human sample comprising 159 immature skulls, between 0 and 8 yr, from the collection of the Strasbourg Medicine Faculty, and on a series of 201 immature *Pan troglodytes* and *Pan paniscus* specimens from the collection of the Royal Museum of Central Africa (Tervuren, Belgium) and from the Museum National d'Histoire Naturelle (Paris, France) (Supplementary Tables S1 and S4).

In contrast with the pattern seen in extant humans and chimpanzees, the Mojokerto tympanic plate is fully ossified, whereas its

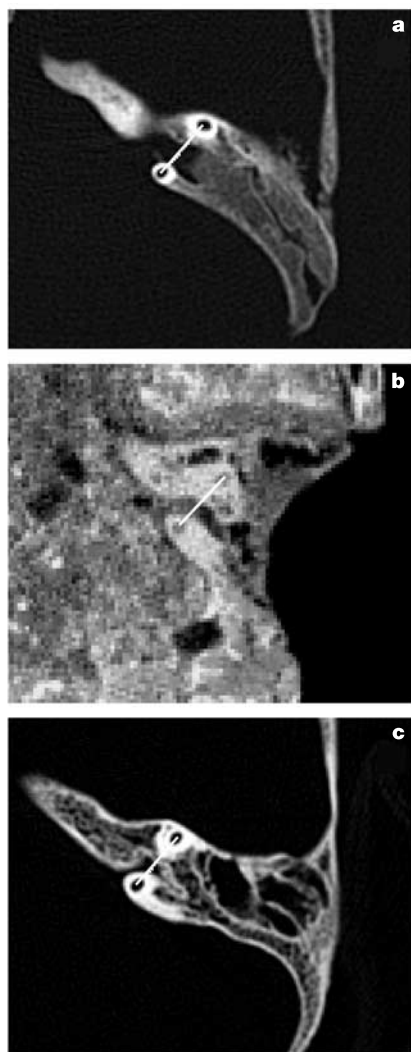


Figure 2 The subarcuate fossa in the right temporal bone of the Mojokerto specimen (**b**) and in two modern specimens (**a**, **c**). The modern specimens are 2 months (**a**) and 18 months (**c**) old. The displayed views correspond to an axial plane parallel to the lateral semicircular canal and are located at the level of maximum development of the fossa. The width of the fossa is measured along a segment joining the two lumens of the anterior semicircular canal (white bar). The exact limits of the fossa in the chosen plane are determined following the method 'HMH' defined by ref. 27. The width is expressed as a percentage (SF) of the length of this segment. Left and right sides were averaged for each individual. The modern specimens were scanned with a Siemens Somatom sensation 16 Scanner (Hôpital de Hautepierre, Strasbourg, France). The scans were made contiguously with a field of view of 60×60 mm (512×512 matrix), a slice thickness of 0.6 mm and a pixel size of 0.117 mm. The skull of the Mojokerto child was scanned with a Mx Twin Scanner (Clinique Pasteur, Toulouse, France). The scans were made contiguously in the transverse plane with a field of view of 204×204 mm ($1,024 \times 1,024$ matrix), a slice thickness of 0.5 mm and a pixel size of 0.199 mm.

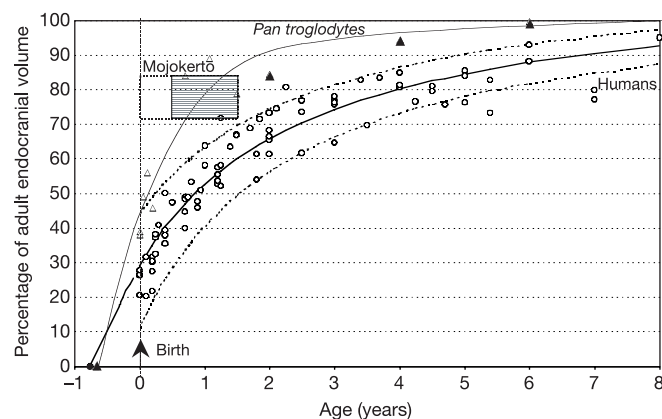


Figure 3 Endocranial volume growth as a percentage of the adult value in Mojokerto, *Pan troglodytes* and extant humans. The Mojokerto specimen is plotted showing the various possibilities in terms of geological and developmental age. The grey area of the Mojokerto box is based only on the estimate from the human model, whereas the white part to the left is based on the estimate from the human and chimpanzee models. The human sample is composed of 93 specimens from the Medicine Faculty of Strasbourg of known calendar age between 0 to 8 yr old. The chimpanzee curve is established from the data from refs 29 and 30 corresponding to a total of 57 specimens between 74 days and 6 yr old (black triangles represent means), and completed by a series of seven individuals of calendar age between birth and 18 months from the Museum National d'Histoire Naturelle (Paris) (white triangles). The endocranial volumes of humans were obtained by direct measurement, and the chimpanzee endocranial volumes either by direct measurements or by medical imaging. The dotted lines correspond to standard errors of estimate around the regression lines (solid lines). The endocranial volume of the Mojokerto specimen was established by medical imaging. On reconstructed sagittal slices separated by 2 mm, the missing portions of the base were reconstructed. The volume was estimated by measuring all the endocranial surfaces with ImageJ software and adding them.

bregmatic area and its subarcuate fossa are still quite immature. The maturation stage of the tympanic plate is not a good age criterion because it can be found open or fully formed in virtually all of the age classes examined (Supplementary Table S5).

The bregmatic area is damaged. Between the anteromedial corners of the parietals a gap of 3.5 mm can either be interpreted as a fontanelle in its final stage of closure or as post-mortem damage. However, even assuming the latter, the parietals are still composed of a single bone table (Fig. 1). In our reference series, an immature pattern of the parietal edges persists only briefly after fontanelle closure, and the development of the diploe near the sagittal suture immediately follows the closure of the anterior fontanelle (Supplementary Fig. S6). In contrast with the modern human condition, the Mojokerto calvaria displays fused hemifrontals with a developed diploe, whereas the parietals are still immature. Similar conditions were met at bregma in our chimpanzee series (Supplementary Fig. S7). The time of closure of the modern human fontanelle was scored by calendar age in our reference series, as well as two other collections: Spitalfields (Natural History Museum, London) and Augier (Musée Orfila, Paris) (Supplementary Tables S2 and S3). In these samples ($n = 297$, between 0 and 5 yr), the closure of the fontanelle occurs between 9 and 34 months.

During infancy and early childhood, the subarcuate fossa is obliterated to form part of the petromastoid canal on the posterior surface of the petrous pyramid. To evaluate the degree of closure of the Mojokerto infant subarcuate fossa, we have measured its width relative to the distance across the anterior semicircular canal (Fig. 2). The bayesian probability of observing in our modern human series a subarcuate fossa similar to that of Mojokerto (SF (see Fig. 2 legend) in the class 20–25%) is $P = 0.20$ between 0–0.5 yr, $P = 0.33$ between 0.5–1 yr and $P = 0.46$ between 1–1.5 yr (Supplementary Table S5). The age determination obtained by analysis of the subarcuate fossa of the Mojokerto specimen is therefore compatible with the lower end of ages estimated from the anterior fontanelle. Taking into account the maturation of the bregmatic area and the subarcuate fossa produces posterior probabilities of calendar ages similar to those obtained with the subarcuate fossa alone (Supplementary Table S5).

If the developmental age of the Mojokerto calvaria is assessed using ape standards, the subarcuate fossa at birth in chimpanzees is generally more closed than it is in the Mojokerto specimen (Supplementary Table S4). The fontanelle is fully closed in all chimpanzee individuals by the age of 3 months and 54% of these individuals also have a fully formed tympanic plate. Combining the results obtained in humans and apes, the most likely age estimate for the Mojokerto specimen is therefore between 0.5 and 1.5 yr, but an age under 0.5 yr cannot be excluded.

A direct measurement of endocranial volume is not possible on the Mojokerto child because the calvaria is filled with matrix. Previous estimates of cranial capacity, using extrapolations from

diameters of the calvaria or direct liquid displacement measurements, ranged from 636 to 730 cubic centimetres^{16,20,21}. CT imaging allowed us to estimate the endocranial volume at 663 cubic centimetres (Fig. 3).

Assuming a geological age for the Mojokerto specimen of approximately 1.8 Myr and a long chronology for the Javan specimens, a comparative sample is provided by early African *Homo erectus* s.l. (KNM-WT 15000 (ref. 22), KNM-ER 3733 (ref. 23) and KNM-ER 3883 (ref. 23)), by the three crania of early *Homo* from Dmanisi (D2280 (ref. 13), D2282 (ref. 13) and D2700 (ref. 14)) and by Sangiran 2 and 4 (Table 1). The endocranial capacity of Mojokerto is about 84% of this adult or sub-adult series. Assuming a short and younger chronological range for the Javan hominids leads to another possible comparison with the series of adult Indonesian *H. erectus* (Sangiran 2, 4, 10, 12, 17, Trinil 2 (ref. 24) and Sangiran IX²⁵) exclusive of the late Ngandong and Sambungmacan specimens. In this case, the cranial capacity of the Mojokerto child is about 72% of the adult size (Fig. 3).

Although our age estimation is primarily based on modern human criteria, the various lines of evidence available on the Mojokerto calvaria more closely match an ape pattern of brain development rate than a modern human one (Fig. 3). On the basis of our modern human sample, the bayesian probability of observing a cranial capacity between 70% and 90% of adult volume is $P = 0.0$ between 0 and 1.0 yr and $P = 0.01$ between 1.0 and 1.5 yr.

Although these results are based on the analysis of only one exceptionally preserved juvenile *H. erectus* skull, they suggest that secondary altriciality was established fairly late in the genus *Homo*, perhaps in the common ancestor of *Homo sapiens* and *Homo neanderthalensis*, which both displayed a very large brain and a reduced pelvic inlet size²⁶. These data also suggest that in *H. erectus* only a short period of brain maturation took place in the extra-maternal environment. This makes it unlikely that early *Homo* had cognitive skills comparable to those of modern humans, and it also implies that complex spoken language emerged relatively late in the course of human evolution. □

Received 8 August 2003; accepted 13 July 2004; doi:10.1038/nature02852.

- Martin, R. D. *Human Brain Evolution in an Ecological Context. Fifty-Second James Arthur Lecture on the Evolution of the Human Brain* (American Museum of Natural History, New York, 1983).
- Dean, C. et al. Growth processes in teeth distinguish modern humans from *Homo erectus* and earlier hominins. *Nature* **414**, 628–631 (2001).
- Rosenberg, K. & Trevathan, W. Birth, obstetrics and human evolution. *Br. J. Obstet. Gynaecol.* **109**, 1199–1206 (2002).
- Gould, S. J. *Ontogeny and Phylogeny* (The Belknap Press of Harvard Univ. Press, Cambridge, Massachusetts, 1977).
- Rosenzweig, M. R. & Bennett, E. L. Psychobiology of plasticity: effects of training and experience on brain and behavior. *Behav. Brain Res.* **78**, 57–65 (1996).
- Huttenlocher, P. R. Morphometric study of human cerebral cortex development. *Neuropsychologia* **28**, 517–527 (1990).
- Ragur, S. Retarded development: the evolutionary mechanism underlying the emergence of the human capacity for language. *J. Mind Behav.* **6**, 451–468 (1985).
- Smith, B. H. & Tompkins, R. L. Toward a life history of the hominidae. *Annu. Rev. Anthropol.* **24**, 257–279 (1995).
- Wood, B. & Collard, M. The human genus. *Science* **284**, 65–71 (1999).
- Walker, A. & Ruff, C. B. in *The Nariokotome Homo erectus Skeleton* (eds Walker, A. & Leakey, R.) 221–233 (Springer, Berlin, 1993).
- Swisher, C. C. III et al. Age of the earliest known hominids in Java, Indonesia. *Science* **263**, 118–121 (1994).
- Huffman, O. Geologic context and age of the Perning/Mojokerto *Homo erectus*, East Java. *J. Hum. Evol.* **40**, 353–362 (2001).
- Gabunia, L. et al. Earliest pleistocene hominid cranial remains from Dmanisi, Republic of Georgia: taxonomy, geological setting, and age. *Science* **288**, 1019–1025 (2000).
- Vekua, A. et al. A new skull of early *Homo* from Dmanisi, Georgia. *Science* **297**, 85–89 (2002).
- von Koenigswald, G. H. R. Ein fossiler hominide aus dem Altpleistocän Ostjavas. *Ingenieur Ned.-Indie* **8**, 149–158 (1936).
- Weinert, H. *Entstehung der Menschenrassen* (Ferdinand Enke, Stuttgart, 1938).
- Weidenreich, F. Foreward to neue *Pithecanthropus*-Funde 1936–1938, ein Beitrag zur Kenntnis der Prahominiden by GHR von Koenigswald. *Wet. Meded.* **28**, 7–14 (1940).
- Jaech, T. in *Paleoanthropology, Morphology and Paleoecology* (ed. Tuttle, R. H.) 311–325 (Mouton, The Hague, 1975).
- Anton, S. C. Developmental age and taxonomic affinity of the Mojokerto child, Java, Indonesia. *Am. J. Phys. Anthropol.* **102**, 497–514 (1997).
- Dubois, E. Racial identity of *Homo soloensis* Oppenoorth (including *Homo modjokertensis* von Koenigswald) and *Sinanthropus pekinensis* Davidson Black. *Konin. Akad. Wet.* **34**, 1180–1185 (1936).

Table 1 Endocranial capacity values for three fossil hominid series

Species	Specimen	Endocranial volume (cc)
Indonesian <i>H. erectus</i>	Sangiran 2	813
	Sangiran 4	908
	Sangiran 10	855
	Sangiran 12	1,059
	Sangiran 17	1,004
	Sangiran IX	850
African early <i>H. erectus</i> s.l.	Trinil 2	940
	KNM-ER 3733	848
	KNM-ER 3883	804
	KNM-WT 15000	880 (adult estimate: 909)
Georgian early <i>Homo</i>	D2280	780
	D2282	650
	D2700	600

cc, cubic centimetres.

21. Riscutia, C. in *Paleoanthropology, Morphology and Paleoecology* (ed. Tuttle, R. H.) 373–375 (Mouton, The Hague, 1975).
22. Begun, D. & Walker, A. in *The Nariokotome Homo erectus Skeleton* (eds Walker, A. & Leakey, R.) 326–358 (Springer, Berlin, 1993).
23. Holloway, R. L. Human paleontological evidence relevant to language behavior. *Hum. Neurobiol.* **2**, 105–114 (1983).
24. Holloway, R. L. The Indonesian *Homo erectus* brain endocasts revisited. *Am. J. Phys. Anthropol.* **55**, 503–521 (1981).
25. Sartono, S. & Tyler, D. E. A New *Homo erectus* Skull from Sangiran, Java: an Announcement 1–4 (Intern. Conf. Human Paleocool., LIPPI, Jakarta, 1993).
26. Rak, Y. & Arensburg, B. Kebara 2 neanderthal pelvis: first look at a complete inlet. *Am. J. Phys. Anthropol.* **73**, 227–231 (1987).
27. Spoor, C. F., Zonneveld, F. W. & Macho, G. A. Linear measurements of cortical bone and dental enamel by computed tomography: applications and problems. *Am. J. Phys. Anthropol.* **91**, 469–484 (1993).
28. Hühne, K. H. *et al.* A new representation of knowledge concerning human anatomy and function. *Nature Med.* **1**, 506–511 (1995).
29. Zuckerman, S. Age-changes in the chimpanzee, with special reference to growth of brain, eruption of teeth, and estimation of age; with a note on the Taung ape. *Proc. Zool. Soc. Lond.* **1**, 1–42 (1928).
30. Schultz, A. H. in *Contributions to Embryology no 170* (ed. Schultz, A. H.) 1–63 (Carnegie Institution of Washington Publication no 518, Washington DC, 1940).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to the following individuals for their assistance in accessing collections and their advice and comments during the preparation of this paper: S. Anton, J. P. Bocquet-Appel, J. Braga, G. Bräuer, M. Braun, P. Darlu, M. Haas, M. von Harling, C. Hemm, J.-L. Kahn, C. Lefèvre, W. van Neer, S. Pääbo, F. Renoult, M. Richards, Ph. Rightmire, F. Schrenk, H. Sick, F. Spoor, T. Striano, J. Treil, W. Wendelen and V. Zeitoun. This research was supported by grants from CNRS and by the Max Planck Society.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.J.H. (hublin@eva.mpg.de).

Genetic evidence supports demic diffusion of Han culture

Bo Wen^{1,2}, Hui Li¹, Daru Lu¹, Xiufeng Song¹, Feng Zhang¹, Yungang He¹, Feng Li¹, Yang Gao¹, Xianyun Mao¹, Liang Zhang¹, Ji Qian¹, Jingze Tan¹, Jianzhong Jin¹, Wei Huang², Ranjan Deka³, Bing Su^{1,3,4}, Ranajit Chakraborty³ & Li Jin^{1,3}

¹State Key Laboratory of Genetic Engineering and Center for Anthropological Studies, School of Life Sciences and Morgan-Tan International Center for Life Sciences, Fudan University, Shanghai 200433, China

²Chinese National Human Genome Center, Shanghai 201203, China

³Center for Genome Information, Department of Environmental Health, University of Cincinnati, Cincinnati, Ohio 45267, USA

⁴Key Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, the Chinese Academy of Sciences, Kunming 650223, China

The spread of culture and language in human populations is explained by two alternative models: the demic diffusion model, which involves mass movement of people; and the cultural diffusion model, which refers to cultural impact between populations and involves limited genetic exchange between them¹. The mechanism of the peopling of Europe has long been debated, a key issue being whether the diffusion of agriculture and language from the Near East was concomitant with a large movement of farmers^{1–3}. Here we show, by systematically analysing Y-chromosome and mitochondrial DNA variation in Han populations, that the pattern of the southward expansion of Han culture is consistent with the demic diffusion model, and that males played a larger role than females in this expansion. The Han people, who all share the same culture and language, exceed 1.16 billion (2000 census), and are by far the largest ethnic group in the world. The

expansion process of Han culture is thus of great interest to researchers in many fields.

According to the historical records, the Hans were descended from the ancient Huaxia tribes of northern China, and the Han culture (that is, the language and its associated cultures) expanded into southern China—the region originally inhabited by the southern natives, including those speaking Daic, Austro-Asiatic and Hmong-Mien languages—in the past two millennia^{4,5}. Studies on classical genetic markers and microsatellites show that the Han people, like East Asians, are divided into two genetically differentiated groups, northern Han and southern Han^{6,8}, separated approximately by the Yangtze river⁹. Differences between these groups in terms of dialect and customs have also been noted¹⁰. Such observations seem to support a mechanism involving primarily cultural diffusion and assimilation (the cultural diffusion model) in Han expansion towards the south. However, the substantial sharing of Y-chromosome and mitochondrial lineages between the two groups^{11,12} and the historical records describing the expansion of Han people⁵ contradict the cultural diffusion model hypothesis of Han expansion. In this study, we aim to examine the alternative hypothesis; that is, that substantial population movements occurred during the expansion of Han culture (the demic diffusion model).

To test this hypothesis, we compared the genetic profiles of southern Hans with their two parental population groups: northern Hans and southern natives, which include the samples of Daic, Hmong-Mien and Austro-Asiatic speaking populations currently residing in China, and in some cases its neighbouring countries. Genetic variation in both the non-recombining region of the Y chromosome (NRY) and mitochondrial DNA (mtDNA)^{13–16} were surveyed in 28 Han populations from most of the provinces in China (see Fig. 1 and Supplementary Table 1 for details).

On the paternal side, southern Hans and northern Hans share similar frequencies of Y-chromosome haplogroups (Supplementary Table 2), which are characterized by two haplogroups carrying the M122-C mutations (O3-M122 and O3e-M134) that are prevalent in almost all Han populations studied (mean and range: 53.8%, 37–71%; 54.2%, 35–74%, for northern and southern Hans, respectively). Haplogroups carrying M119-C (O1* and O1b) and/or M95-T (O2a* and O2a1) (following the nomenclature of the Y Chromosome Consortium) which are prevalent in southern natives, are more frequent in southern Hans (19%, 3–42%) than in northern Hans (5%, 1–10%). In addition, haplogroups O1b-M110, O2a1-M88 and O3d-M7, which are prevalent in southern natives¹⁷, were only observed in some southern Hans (4% on average), but not in northern Hans. Therefore, the contribution of southern natives in southern Hans is limited, if we assume that the frequency distribution of Y lineages in southern natives represents that before the expansion of Han culture that started 2,000 yr ago⁵. The results of analysis of molecular variance (AMOVA) further indicate that northern Hans and southern Hans are not significantly different in their Y haplogroups ($F_{ST} = 0.006$, $P > 0.05$), demonstrating that southern Hans bear a high resemblance to northern Hans in their male lineages.

On the maternal side, however, the mtDNA haplogroup distribution showed substantial differentiation between northern Hans and southern Hans (Supplementary Table 3). The overall frequencies of the northern East Asian-dominating haplogroups (A, C, D, G, M8a, Y and Z) are much higher in northern Hans (55%, 49–64%) than are those in southern Hans (36%, 19–52%). In contrast, the frequency of the haplogroups that are dominant lineages (B, F, R9a, R9b and N9a) in southern natives^{12,14,18} is much higher in southern (55%, 36–72%) than it is in northern Hans (33%, 18–42%). Northern and southern Hans are significantly different in their mtDNA lineages ($F_{ST} = 0.006$, $P < 10^{-5}$). Although the F_{ST} values between northern and southern Hans are similar for mtDNA and the Y chromosome, F_{ST} accounts for 56% of the total among-

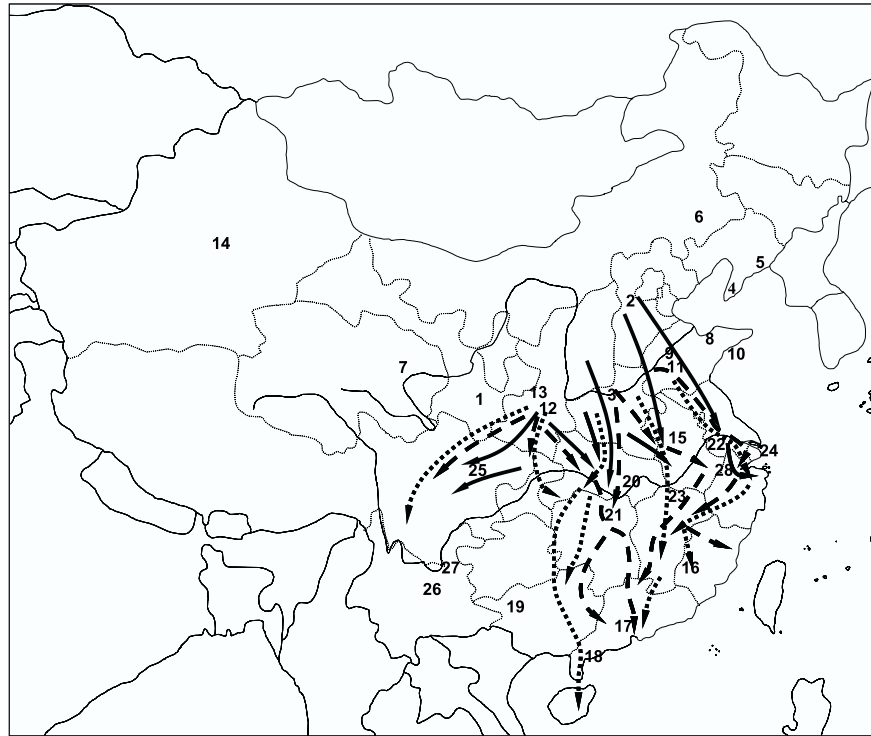


Figure 1 Geographic distribution of sampled populations. Shown are the three waves of north-to-south migrations according to historical record. The identifications of populations are given in Supplementary Table 1. Populations 1–14 are northern Hans, and 15–28 are southern Hans. The solid, dashed and dotted arrows refer to the first, second and third waves of migrations, respectively. The first wave involving 0.9 million (approximately

one-sixth of the southern population at that time) occurred during the Western Jin Dynasty (AD 265–316); the second migration, more extensive than the first, took place during the Tang Dynasty (AD 618–907); and the third wave, including ~5 million immigrants, occurred during the Southern Song Dynasty (AD 1127–1279).

population variation for mtDNA but only accounts for 18% for the Y chromosome.

A principal component analysis is consistent with the observation based on the distribution of the haplogroups in Han populations. For the NRY, almost all Han populations cluster together in the upper right-hand part of Fig. 2a. Northern Hans and southern natives are separated by the second principal component (PC2) and southern Hans' PC2 values lie between northern Hans and southern natives but are much closer to northern Hans (northern Han, 0.58 ± 0.01 ; southern Han, 0.46 ± 0.03 ; southern native, -0.32 ± 0.05), implying that the southern Hans are paternally similar to northern Hans, with limited influence from southern natives. In contrast, for mtDNA, northern Hans and southern natives are distinctly separated by PC2 (Fig. 2b), and southern Hans are located between them but are closer to southern natives (northern Han, 0.56 ± 0.02 ; southern Han, 0.09 ± 0.06 ; southern native, -0.23 ± 0.04), indicating a much more substantial admixture in southern Hans' female gene pool than in its male counterpart.

The relative contribution of the two parental populations (northern Hans and southern natives) in southern Hans was estimated by two different statistics^{19,20}, which are less biased than other statistics for single-locus data²¹ (Table 1). The estimations of the admixture coefficient (M , proportion of northern Han contribution) from the two methods are highly consistent (for the Y chromosome, $r = 0.922$, $P < 0.01$; for mtDNA, $r = 0.970$, $P < 0.01$). For the Y chromosome, all southern Hans showed a high proportion of northern Han contribution (M_{BE} : 0.82 ± 0.14 , range from 0.54 to 1; M_{RH} : 0.82 ± 0.12 , range from 0.61 to 0.97) (see refs 20 and 19 for definitions of M_{BE} and M_{RH} , respectively) indicating that males from the northern Hans are the primary contributor to the gene pool of the southern Hans. In contrast, northern Hans and southern natives contributed almost equally to the southern Hans' mtDNA

Table 1 Northern Han admixture proportion in southern Hans

Population	Y Chromosome		mtDNA	
	$M_{BE} (\pm s.e.m)$	M_{RH}	$M_{BE} (\pm s.e.m)$	M_{RH}
Anhui	0.868 ± 0.119	0.929	0.816 ± 0.214	0.755
Fujian	1	0.966	0.341 ± 0.206	0.248
Guangdong1	0.677 ± 0.121	0.669	0.149 ± 0.181	0.068
Guangdong2	ND	ND	0.298 ± 0.247	0.312
Guangxi	0.543 ± 0.174	0.608	0.451 ± 0.263	0.249
Hubei	0.981 ± 0.122	0.949	0.946 ± 0.261	0.907
Hunan	0.732 ± 0.219	0.657	0.565 ± 0.297	0.490
Jiangsu	0.789 ± 0.078	0.821	0.811 ± 0.177	0.786
Jiangxi	0.804 ± 0.113	0.829	0.374 ± 0.343	0.424
Shanghai	0.819 ± 0.087	0.902	0.845 ± 0.179	0.833
Sichuan	0.750 ± 0.118	0.713	0.509 ± 0.166	0.498
Yunnan1	1	0.915	0.376 ± 0.221	0.245
Yunnan2	0.935 ± 0.088	0.924	0.733 ± 0.192	0.645
Zhejiang	0.751 ± 0.084	0.763	0.631 ± 0.180	0.540
Average	0.819	0.819	0.560	0.500

M_{BE} and M_{RH} refer to the statistics described in refs 20 and 19, respectively. The standard error of M_{BE} was obtained by bootstrap with 1,000 replications. The proportions of contribution from northern Hans were estimated using northern Hans and southern natives as the parental populations of the southern Hans. It was assumed that the allele frequency in the southern natives remained unchanged before and after the admixture, which started about 2,000 yr ago, and the genetic exchange between northern Hans and southern natives has been limited. In fact, the gene flow from northern Hans to southern natives has been larger than that from southern natives to northern Hans; therefore, the level of admixture presented in this table is underestimated and is without proper adjustment. The demic expansion of Han would have been more pronounced than was observed in this study.

gene pool (M_{BE} : 0.56 ± 0.24 [0.15, 0.95]; M_{RH} : 0.50 ± 0.26 [0.07, 0.91]). The contribution of northern Hans to southern Hans is significantly higher in the paternal lineage than in the maternal lineage collectively (t -test, $P < 0.01$) or individually (11 out of 13 populations for M_{BE} , and 13 out of 13 populations for M_{RH} ; $P < 0.01$, assuming a null binomial distribution with equal male and female contributions), indicating a strong sex-biased popu-

lation admixture in southern Hans. The proportions of northern Han contribution (M) in southern Hans showed a clinal geographic pattern, which decreases from north to south. The M s in southern Hans are positively correlated with latitude ($r^2 = 0.569$, $P < 0.01$) for mtDNA, but are not significant for the Y chromosome ($r^2 = 0.072$, $P > 0.05$), because the difference of M s in the paternal lineage among southern Hans is too small to create a statistically significant trend.

We provide two lines of evidence supporting the demic diffusion hypothesis for the expansion of Han culture. First, almost all Han populations bear a high resemblance in Y-chromosome haplogroup distribution, and the result of principal component analysis indicated that almost all Han populations form a tight cluster in their Y chromosome. Second, the estimated contribution of northern Hans to southern Hans is substantial in both paternal and maternal lineages and a geographic cline exists for mtDNA. It is noteworthy that the expansion process was dominated by males, as is shown by a greater contribution to the Y-chromosome than the mtDNA from northern Hans to southern Hans. A sex-biased admixture pattern was also observed in Tibeto-Burman-speaking populations²².

According to the historical records, there were continuous southward movements of Han people due to warfare and famine in the north, as illustrated by three waves of large-scale migrations (Fig. 1). Aside from these three waves, other smaller southward migrations also occurred during almost all periods in the past two millennia.

Our genetic observation is thus in line with the historical accounts. The massive movement of the northern immigrants led to a change in genetic makeup in southern China, and resulted in the demographic expansion of Han people as well as their culture. Except for these massive population movements, gene flow between northern Hans, southern Hans and southern natives also contributed to the admixture which shaped the genetic profile of the extant populations. □

Methods

Samples

Blood samples of 871 unrelated anonymous individuals from 17 Han populations were collected across China. Genomic DNA was extracted by the phenol-chloroform method. By integrating the additional data obtained from the literatures on the Y chromosome and on mtDNA variation, the final sample sizes for analysis expanded to 1,289 individuals (23 Han populations) for the Y chromosome and 1,119 individuals (23 Han populations) for mtDNA. These samples encompass most of the provinces in China (Fig. 1 and Supplementary Table 1).

Genetic markers

Thirteen bi-allelic Y-chromosome markers, YAP, M15, M130, M89, M9, M122, M134, M119, M110, M95, M88, M45 and M120 were typed by polymerase chain reaction-restriction-fragment length polymorphism methods¹¹. These markers are highly informative in East Asians²³ and define 13 haplogroups following the Y Chromosome Consortium nomenclature²⁴.

The HVS-1 of mtDNA and eight coding region variations, 9-bp deletion, 10397 *AluI*, 5176 *AluI*, 4831 *HhaI*, 13259 *HincII*, 663 *HaeIII*, 12406 *HpaI* and 9820 *HinfI* were sequenced and genotyped as in our previous report²². Both the HVS-1 motif and the coding region variations were used to infer haplogroups following the phylogeny of East Asian mtDNAs¹⁸.

Data analysis

Population relationship was investigated by principal component analysis, which was conducted using mtDNA and Y-chromosome haplogroup frequencies and SPSS10.0 software (SPSS Inc.). The genetic difference between northern and southern Hans was tested by AMOVA²⁵, using ARLEQUIN software²⁶. ADMIX 2.0 (ref. 27) and LEADMIX²¹ software were used to estimate the level of admixture of the northern Hans and southern natives in the southern Han populations, using two different statistics^{19–20}. The selection of parental populations is critical for appropriate estimation of admixture proportion and we were careful to minimize bias by using large data sets across East Asia. In this analysis, the average haplogroup frequencies (for Y-chromosome or mtDNA markers, respectively) of northern Hans (arithmetic mean of 10 northern Hans) were taken for the northern parental population. The frequency of southern natives was estimated by the average of three groups including Austro-Asiatic (NRY, 6 populations; mtDNA, 5 populations), Daic (NRY, 22 populations; mtDNA, 11 populations) and Hmong-Mien (NRY, 18 populations; mtDNA, 14 populations). The geographic pattern of Han populations was revealed by the linear regression analysis of admixture proportion against the latitudes of samples^{1,3}.

Received 28 April; accepted 20 July 2004; doi:10.1038/nature02878.

1. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* (Princeton Univ. Press, Princeton, 1994).
2. Sokal, R. R., Oden, N. L. & Wilson, C. Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature* **351**, 143–145 (1991).
3. Chikhi, L. *et al.* Y genetic data support the Neolithic demic diffusion model. *Proc. Natl Acad. Sci. USA* **99**, 11008–11013 (2002).
4. Fei, X. T. *The Pattern of Diversity in Unity of the Chinese Nation* (Central Univ. for Nationalities Press, Beijing, 1999).
5. Ge, J. X., Wu, S. D. & Chao, S. J. *Zhongguo yimin shi (The Migration History of China)* (Fujian People's Publishing House, Fuzhou, China, 1997).
6. Zhao, T. M. & Lee, T. D. Gm and Km allotypes in 74 Chinese populations: a hypothesis of the origin of the Chinese nation. *Hum. Genet.* **83**, 101–110 (1989).
7. Du, R. F., Xiao, C. J. & Cavalli-Sforza, L. L. Genetic distances calculated on gene frequencies of 38 loci. *Sci. China* **40**, 613 (1997).
8. Chu, J. Y. *et al.* Genetic relationship of populations in China. *Proc. Natl Acad. Sci. USA* **95**, 11763–11768 (1998).
9. Xiao, C. J. *et al.* Principal component analysis of gene frequencies of Chinese populations. *Sci. China* **43**, 472–481 (2000).
10. Xu, Y. T. A brief study on the origin of Han nationality. *J. Centr. Univ. Natl* **30**, 59–64 (2003).
11. Su, B. *et al.* Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Hum. Genet.* **107**, 582–590 (2000).
12. Yao, Y. G. *et al.* Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am. J. Hum. Genet.* **70**, 635–651 (2002).
13. Cavalli-Sforza, L. L. & Feldman, M. W. The application of molecular genetic approaches to the study of human evolution. *Nature Genet.* **33**, 266–275 (2003).
14. Wallace, D. C., Brown, M. D. & Lott, M. T. Nucleotide mitochondrial DNA variation in human evolution and disease. *Gene* **238**, 211–230 (1999).
15. Underhill, P. A. *et al.* Y chromosome sequence variation and the history of human populations. *Nature Genet.* **26**, 358–361 (2000).
16. Jobling, M. A. & Tyler-Smith, C. The human Y chromosome: an evolutionary marker comes of age.

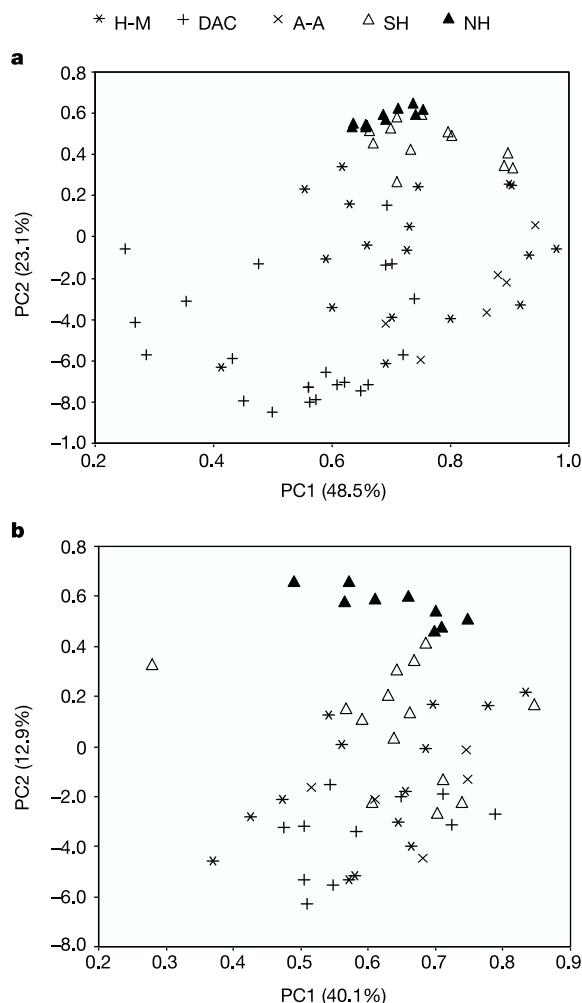


Figure 2 Principal component plot. **a**, **b**, Plots of Y-chromosome (**a**) and mtDNA (**b**) haplogroup frequency. Population groups: H-M, Hmong-Mien; DAC, Daic; A-A, Austro-Asiatic; SH, southern Han; NH, northern Han.

Nature Rev. Genet. 4, 598–612 (2003).

17. Su, B. *et al.* Y-chromosome evidence for a northward migration of modern humans into eastern Asia during the last ice age. *Am. J. Hum. Genet.* **65**, 1718–1724 (1999).
18. Kivisild, T. *et al.* The emerging limbs and twigs of the East Asian mtDNA tree. *Mol. Biol. Evol.* **19**, 1737–1751 (2002).
19. Roberts, D. F. & Hiorns, R. W. Methods of analysis of the genetic composition of a hybrid population. *Hum. Biol.* **37**, 38–43 (1965).
20. Bertorelle, G. & Excoffier, L. Inferring admixture proportions from molecular data. *Mol. Biol. Evol.* **15**, 1298–1311 (1998).
21. Wang, J. Maximum-likelihood estimation of admixture proportions from genetic data. *Genetics* **164**, 747–765 (2003).
22. Wen, B. *et al.* Analyses of genetic structure of Tibeto-Burman populations revealed a gender-biased admixture in southern Tibeto-Burmans. *Am. J. Hum. Genet.* **74**, 856–865 (2004).
23. Jin, L. & Su, B. Natives or immigrants: modern human origin in East Asia. *Nature Rev. Genet.* **1**, 126–133 (2000).
24. The Y Chromosome Consortium, A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* **12**, 339–348 (2002).
25. Excoffier, L., Smouse, P. E. & Quattro, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491 (1992).
26. Schneider, S., *et al.* Arlequin: Ver. 2.000. A software for population genetic analysis. (Genetics and Biometry Laboratory, Univ. of Geneva, Geneva, 2000).
27. Dupanloup, I. & Bertorelle, G. Inferring admixture proportions from molecular data: extension to any number of parental populations. *Mol. Biol. Evol.* **18**, 672–675 (2001).
28. Chakraborty, R. Gene admixture in human populations: Models and predictions. *Yb. Phys. Anthropol.* **29**, 1–43 (1986).
29. Sans, M. *et al.* Unequal contributions of male and female gene pools from parental populations in the African descendants of the city of Melo, Uruguay. *Am. J. Phys. Anthropol.* **118**, 33–44 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank all of the donors for making this work possible. The data collection was supported by NSFC and STCSM to Fudan and a NSF grant to L.J. L.J., R.D. and R.C. are supported by NIH.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.J. (lijin@fudan.edu.cn or lijin@uc.edu). The mtDNA HVS-1 sequences of 711 individuals from 15 Han populations were submitted to GenBank with accession numbers AY594701–AY595411.

Post-mating clutch piracy in an amphibian

David R. Vieites^{1,2*}, Sandra Nieto-Román^{1,3}, Marta Barluenga², Antonio Palanca¹, Miguel Vences³ & Axel Meyer²

¹Laboratorio de Anatomía Animal, Departamento de Ecología e Biología Animal, Facultad de Ciencias Biológicas, Universidade de Vigo, Buzón 137, 36201 Vigo, Spain

²Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, 78457 Konstanz, Germany

³Institute for Biodiversity and Ecosystem Dynamics and Zoological Museum, University of Amsterdam, Mauritskade 61, 1092 AD Amsterdam, The Netherlands

* Present address: Museum of Vertebrate Zoology and Department of Integrative Biology, 3101 Valley Life Sciences Building, University of California, Berkeley, California 94720-3160, USA

Female multiple mating and alternative mating systems can decrease the opportunity for sexual selection^{1–3}. Sperm competition is often the outcome of females mating with multiple males and has been observed in many animals^{1,4–7}, and alternative reproductive systems are widespread among species with external fertilization and parental care^{3,8–10}. Multiple paternity without associated complex behaviour related to mating or parental care is also seen in simultaneously spawning amphibians^{11–15} and fishes¹⁶ that release gametes into water. Here we report ‘clutch piracy’ in a montane population of the common frog *Rana temporaria*, a reproductive behaviour previously unknown in vertebrates with external fertilization. Males of this species clasp the females and the pair deposits one spherical clutch of eggs. No

parental care is provided. ‘Pirate’ males search for freshly laid clutches, clasp them as they would do a female and fertilize the eggs that were left unfertilized by the ‘parental’ male. This behaviour does not seem to be size-dependent, and some males mate with a female and perform clutch piracy in the same season. Piracy affected 84% of the clutches and in some cases increased the proportion of eggs fertilized, providing direct fitness benefits both for the pirate males and the females¹⁷. Sexual selection—probably caused by a strong male-biased sex ratio—occurs in this population, as indicated by size-assortative mating; however, clutch piracy may reduce its impact. This provides a good model to explore how alternative mating strategies can affect the intensity of sexual selection.

Anuran amphibians have a wide diversity of reproductive modes, but external aquatic fertilization without parental care is the ancestral and most widespread strategy¹⁸. Only a few instances of multiple paternity have been demonstrated in frogs and those were considered to be the result of polyandrous matings, in which several males mate simultaneously with a female^{11,13,14}. In the common frog *R. temporaria*, one of the most widespread Palearctic amphibians¹⁹, multiple paternity has been detected through allozyme analyses of tadpole kin groups, and was interpreted as being the consequence of high concentrations of spermatozoa in the water during simultaneous spawning¹².

R. temporaria is an explosive pond breeder that often reproduces immediately after the melting of the ice cover. Breeding is usually nocturnal^{12,20} and males form large breeding aggregations. We monitored a high altitude population of common frogs in a medium-sized pond (540 m²) during three consecutive breeding periods (2001–2003) in the central Pyrenean mountains, Spain (42°49′ N, 0°17′ W, about 2200 m above sea level). Breeding was exclusively diurnal due to low temperatures at night (Supplementary Information A), which permitted us to conduct detailed behavioural observations in the field and to measure and mark most individuals in this population. Males aggregated in a specific area of the pond, where clutches were also laid. Male density at the

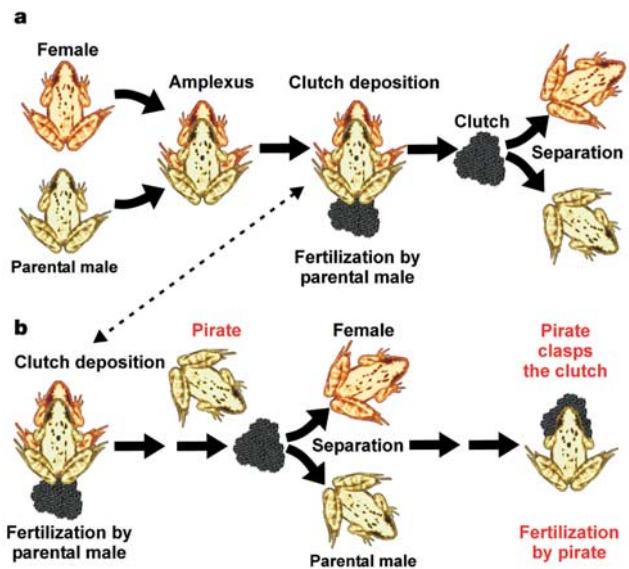


Figure 1 Schematic representation of mating systems in *R. temporaria*. **a**, Females arrive at the breeding ponds and are clasped in the axillary region (‘amplexus’) by a male (the ‘parental’ male). The female deposits a single, spherical clutch of eggs. The parental male simultaneously releases his sperm and thereby fertilizes the eggs externally. Subsequently both parents leave the clutch. **b**, ‘Pirate’ males search for freshly laid clutches, clasp them and release their sperm, sometimes crawling into the clutch to gain access to the internal eggs.

spawning site ranged from 0.40–4.17 males per m^2 (mean $\pm$ s.d. = 1.99 ± 1.02 males per m^2) during the reproductive period. Females arriving at the pond were approached and sometimes clasped by several males, but they only spawned while in amplexus with a single male (Fig. 1). Often hours or even days were spent before eggs were deposited at a calm site in the pond. Altogether, 119 females spawned in 2001, and the population of males was estimated at about 250 in that year. Males were slightly smaller than females (snout–vent length 54–76.5 mm, mean $\pm$ s.d. = 65.8 ± 4.6 mm, versus 58.5–78.5 mm, mean $\pm$ s.d. = 68.7 ± 5.2 mm, respectively; $U_{44,19} = 276.5$, $P = 0.034$).

One or several sneaking males often followed pairs in amplexus, during the pair's search for an appropriate egg deposition site. After eggs had been deposited and the parents had left, these individuals frequently clasped the clutch and released their sperm (Figs 1 and 2; Supplementary Video). This type of behaviour is here termed 'clutch piracy'. In one case a clutch was clasped by a pirate and torn away from the parents at the moment of spawning. In other instances piracy was performed by single males, or by groups of males actively searching the pond for freshly laid clutches. In numerous cases the pirates crawled into the spherical clutch, thereby accessing the eggs in its interior. In 2001, piracy was recorded for 84% of the 119 clutches, by 1–16 males per clutch (mean $\pm$ s.d. = 5 ± 4 males). In 45% of the 119 clutches the pirates gained access to the interior eggs. They spent 35–387 s fertilizing a clutch (mean $\pm$ s.d. = 96.4 ± 105 s). Piracy usually took place within seconds after the clutches were laid but was also observed for up to 2 h later. *U*-tests of parentals versus pirates, parentals versus all males, and pirates versus all males found no significant size differences among parental males (mean snout–vent length $\pm$ s.d. = 66.1 ± 4.2 mm, $n = 27$) and pirates (mean $\pm$ s.d. = 65.8 ± 4.9 mm, $n = 28$), or the total male population (mean $\pm$ s.d. = 66.3 ± 5.6 mm, $n = 230$) ($P > 0.05$ in all cases). On one occasion, a male performed clutch piracy less than 3 h after

having previously acted as a parental male in amplexus with a female. Hence, pirate males in *R. temporaria* are not characterized by a fixed behavioural role, or by a specific age or body size. This differs from examples in other species (mainly fish), in which large bourgeois males coexist with smaller sneakers or satellites^{8,9}.

To determine the proportion of eggs fertilized by pirates we sampled 16 pirated clutches after monitoring them for 1 h after spawning. Subsequently the parentals and pirates were captured and tissue samples were taken by toe clipping. Molecular paternity analyses detected successful fertilization by pirate males in seven of these clutches (mean $\pm$ s.d. = $26.1 \pm 36.2\%$ of the eggs). Of the total 319 embryos analysed, 77 (24.1%) resulted from fertilization by pirate males. The highest paternity (100%) of a pirate was recorded in a clutch that had been pirated 1 min after spawning. The second highest pirate paternity (95%) was found in a clutch that had been clasped and carried away by the pirate while the parental male was still releasing his sperm. In four out of the seven pirated clutches, the full pirate paternity could be assigned to the first pirate arriving at the clutch. In the remaining three clutches, another or several pirates succeeded in fertilizing parts of the clutch.

Multiple paternity in frogs has so far usually been attributed to group spawning and simultaneous mating with several males^{11,14,21}, which can be associated with a reduced overall success of fertilization¹³. Considerable variation in the number of fertilized eggs per clutch has been reported previously for *R. temporaria* (5–100%)^{22,23}. By experimentally rearing 25 pirated and 31 non-pirated clutches we found that under specific circumstances clutch piracy helps to increase the number of eggs fertilized per clutch (Fig. 3).

Alternative mating strategies are expected to evolve in the presence of strong sexual selection or whenever some males are excluded from mating¹. Both conditions might apply to the studied population of *R. temporaria* in the Pyrenees. Only a fraction of all females arrived each day at the pond, significantly skewing the operational sex ratio (OSR) towards a strong male bias (daily OSR: 0.07–0.20 gravid females per sexually active male; mean $\pm$ s.d. = 0.13 ± 0.04). Using a conservative approach—that is, assuming that each parental male mated with only one female—the estimated opportunity for sexual selection (the I_{mates} parameter)¹ is higher ($I_{\text{mates}} = 1.28$, see Methods) than would be expected if mating were random ($I_{\text{mates}} = 1$). Size-assortative mating provides further evidence for the existence of sexual selection in this population. Parental male and female sizes were significantly

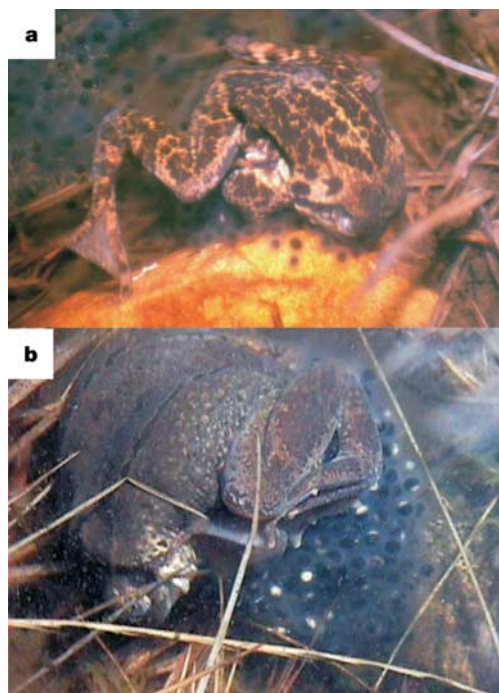


Figure 2 Images of clutch piracy in *R. temporaria*. The pirate male clasps a freshly laid clutch and coils his body over the clutch while fertilizing it. **a**, A male has crawled into the clutch to gain access to the interior eggs. **b**, The male has clasped the clutch at the moment of spawning, while the parental male is still releasing his sperm; in this case, the pirate male succeeded to fertilize 95% of the genotyped eggs.

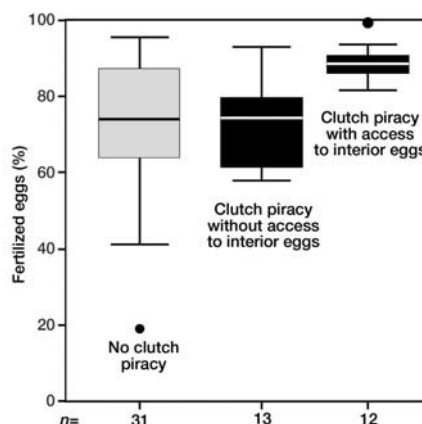


Figure 3 Percentage of fertilized eggs in clutches not exposed (grey) and exposed (black) to clutch piracy. The percentage of eggs of a clutch fertilized in the absence of piracy was 19.2–95.7% (mean $\pm$ s.d. = $72.9 \pm 18.0\%$). This percentage was not significantly different to that found for all clutches exposed to piracy ($U_{31,25} = 308$, $P = 0.19$), but in clutches in which the pirate had access to the interior eggs the percentage was higher ($U_{31,12} = 80$, $P = 0.003$). Error bars are 5% and 95% percentiles; 25% and 75% percentiles delimit the box, and the median is shown as a line. Dots represent outliers.

correlated in 2001 (Nonparametric rank correlation, Spearman's correlation coefficient $\rho = 0.61$, $P = 0.0032$, $n = 21$) and over the whole three-year period ($\rho = 0.49$, $P = 0.0004$, $n = 48$, see Supplementary Information D), although males in *R. temporaria* are known to be unselective regarding female size²⁴. These results support the hypothesis that clutch piracy evolved in the presence, and probably as a consequence, of the highly male-biased OSR observed in this population.

Clutch piracy can provide direct fitness advantages for individual male and female frogs. Pirates gain the opportunity to fertilize eggs when female availability is low, and females gain offspring due to clutch piracy. Producing offspring sired by different fathers has the further advantages of increasing the genetic diversity and therefore the fitness of the progeny, and reduces potential genetic incompatibilities with individual males^{7,17,25}. The effects of clutch piracy on the fitness of parental males is uncertain, except when pirates are able to steal a clutch before it is fertilized by the parent. Males increase their probability of reproduction in the presence of strong selection. This implies that males using this alternative mating behaviour had a fitness exceeding that of males using only the conventional behaviour when invading the population, and have a fitness at least equal to that of conventional males to persist in this population. Estimating the average proportion of time or energy used by the males performing either behaviour would allow these predictions to be rigorously tested¹.

Clutch piracy could result in a highly polygynandrous genetic mating system³, because males can fertilize eggs in several clutches and eggs in one single clutch can be fertilized by several males. We are not aware of any other instance of multiple paternity caused by similar post-mating behaviours in animals with external fertilization. Egg piracy is known in fishes^{26,27} but refers to situations in which eggs are fertilized by other males, then stolen by the pirate and stored in his nest, possibly to attract females, distract predators or serve as food³. Brood parasitism usually involves of furtive participation in fertilization during the spawning procedure, but not the active search for clutches and the fertilization of yet unfertilized eggs after the completion of mating. Our paternity analysis gives results comparable to previous indirect studies¹², suggesting that clutch piracy, although not reported before, could be frequent throughout the distribution area of *R. temporaria*. This qualifies the species to be used as a model system for testing further predictions of sexual selection theory in externally fertilizing vertebrates. □

Methods

Clutch sampling

For molecular paternity analyses, a minimum of 15 eggs per clutch (about 4–5% of the total number of eggs) were randomly collected from the upper part, lower part and interior of clutches. Tissue samples were preserved in 99% ethanol. Data from a total of 319 eggs from 16 clutches (14% of the total; 19 ± 4 eggs per clutch) were available for molecular paternity analyses.

Behavioural observations and clutch rearing

Reproductive behaviour was recorded during the whole breeding period of 2001 (2 weeks of permanent observations between 8:00 and 19:00 hours; Supplementary Information A). Behavioural patterns were documented through photography and video. Daily numbers of males, females, couples in amplexus and clutches laid, and the time of spawning were recorded. The daily OSR was calculated as the relative number of females with fertilizable oocytes per sexually active male²⁰. To estimate the opportunity for sexual selection, we calculated the I_{males} parameter¹ as the difference in the opportunity for selection between males (I_{males}) and females (I_{females}), minus $(\text{OSR} - 1) \times I_{\text{females}}$ (see Supplementary Information B). Fieldwork in 2002 followed the scheme of the previous year but was limited to observations to avoid disturbing the population. For each clutch we counted the number of pirate males attempting to fertilize it, the time that they spent in contact with the clutch (calculated for the whole group in cases of simultaneous piracy by larger numbers of pirates) and whether they gained access to the interior of the clutch. One hour after they had been laid, 25 of these clutches and 31 control clutches that had been isolated from pirates immediately after spawning were taken to the laboratory. Eggs were reared to developmental Gosner stage 25 in small aquaria ($22 \times 18 \times 15$ cm) inside the laboratory, to preserve them from direct UV-B radiation that could be a cause of egg mortality. The proportion of fertilized eggs was recorded for each clutch by counting the number of developing embryos that reached the earlier Gosner stages. Every 2 days, part of the water

was removed and replaced by clean water from the study pond. The light:dark period in the laboratory was 17:7, similar to field conditions.

Microsatellite paternity analysis

All DNA was isolated from muscle in adults or whole embryos in clutches. We tested a total of ten microsatellites (*RtU4* and *RtU7*; *Rtempu1*, *Rtempu4*, *Rtempu5* and *Rtempu6*; *Rtub* and *Rtuo*; and *RECALQ* and *RC08604*; ref. 28); five of these (*RtU4*, *RtU7*, *Rtempu4*, *Rtempu6* and *Rtub*; Supplementary Information C) were successfully amplified by polymerase chain reaction in all individuals and showed a sufficient degree of polymorphism to determine paternity unambiguously. We analysed fragment length with Genotyper 3.7 (Applied Biosystems). Observed and expected heterozygosity and allele frequency for each locus were calculated for the complete adult population with Cervus 2.0 (ref. 29). The probability of detecting multiple mating (P_{DM}) increases with the number of loci, locus polymorphism, number of offspring analysed per clutch and number of specimens genetically contributing to clutch paternity³⁰. We used allele frequency data to calculate this variable in our population using the *PrDM* software³⁰. This program calculates the P_{DM} using the following: (1) number of loci; (2) the number of alleles at each locus and their population frequencies; (3) the number of sires contributing to a multiple mated clutch; and (4) their reproductive skew. Although the number of observed potential sires in our population ranged from 1 to 17, we used a conservative model of only two sires; one with the mean parental proportion of eggs fertilized (0.73) and the other with the remainder (0.27). The genotype of the parents was not specified in order to calculate the mean P_{DM} for the population³⁰. We performed ten replicates of the analysis for 15 offspring samples and calculated the mean P_{DM} . The mean P_{DM} was high ($0.981 \pm 4 \times 10^{-4}$), indicating that the five highly polymorphic loci and samples analysed were sufficient to detect multiple paternity in our population. Paternity of offspring was assigned by direct comparison between their alleles and the alleles present in the parental and pirate males in the five loci analysed. Offspring that were homozygous for at least one allele present in the mother but not in the parental male, or with at least one allele not present in either of the parentals for the five loci analysed, were considered as offspring sired by pirate males.

Received 20 April; accepted 22 July 2004; doi:10.1038/nature02879.

- Shuster, S. M. & Wade, M. J. *Mating Systems and Strategies. Monographs in Behavior and Ecology* (Princeton Univ. Press, New Jersey, 2003).
- Jones, A. G., Walker, D., Kvarnemo, C., Lindstrom, K. & Avise, J. C. How cuckoldry can decrease the opportunity for sexual selection: Data and theory from a genetic parentage analysis of the sand goby, *Pomatoschistus minutus*. *Proc. Natl Acad. Sci. USA* **98**, 9151–9156 (2001).
- Avise, J. C. *et al.* Genetic mating systems and reproductive natural histories of fishes: lessons for ecology and evolution. *Annu. Rev. Genet.* **36**, 19–45 (2002).
- Birkhead, T. R. & Parker, G. A. in *Behavioural Ecology: An Evolutionary Approach* (eds Krebs, J. R. & Davies, N. B.) 121–148 (Blackwell, Oxford, 1997).
- Olsson, M. & Madsen, T. Promiscuity in sand lizards (*Lacerta agilis*) and adder snakes (*Vipera berus*): causes and consequences. *J. Hered.* **92**, 190–197 (2001).
- Pearse, D. E. & Avise, J. C. Turtle mating systems: behavior, sperm storage, and genetic paternity. *J. Hered.* **92**, 206–211 (2001).
- Garner, T. W. J. & Schmidt, B. R. Relatedness, body size and paternity in the alpine newt, *Triturus alpestris*. *Proc. R. Soc. Lond. B* **270**, 619–624 (2003).
- Neff, B. D. Stabilizing selection on genomic divergence in a wild fish population. *Proc. Natl Acad. Sci. USA* **101**, 2381–2385 (2004).
- Taborsky, M. The evolution of bourgeois, parasitic, and cooperative reproductive behaviors in fishes. *J. Hered.* **92**, 100–110 (2001).
- Clutton-Brock, T. H. *The Evolution of Parental Care. Monographs in Behavior and Ecology* (Princeton Univ. Press, New Jersey, 1991).
- D'Ooge, C. A. & Turner, B. J. Multiple paternity in the red-eye treefrog *Agalychnis callidryas* (Cope). *Mol. Ecol.* **4**, 505–508 (1995).
- Laurila, A. & Seppä, P. Multiple paternity in the common frog (*Rana temporaria*): genetic evidence from tadpole kin groups. *Biol. J. Linn. Soc.* **63**, 221–232 (1998).
- Byrne, P. G. & Roberts, J. D. Simultaneous mating with multiple males reduces fertilization success in the myobatrachid frog *Crinia georgiana*. *Proc. R. Soc. Lond. B* **266**, 717–721 (1999).
- Roberts, J. D., Standish, R. J., Byrne, P. G. & Doughty, P. Synchronous polyandry and multiple paternity in the frog *Crinia georgiana* (Anura: Myobatrachidae). *Anim. Behav.* **57**, 721–726 (1999).
- Lodé, T. & Lesbarrères, D. Multiple paternity in *Rana dalmatina*, a monogamous territorial breeding anuran. *Naturwissenschaften* **91**, 44–47 (2004).
- Bekkevold, D., Hansen, M. M. & Loeschcke, V. Male reproductive competition in spawning aggregations of cod (*Gadus morhua*, L.). *Mol. Ecol.* **11**, 91–102 (2002).
- Jennions, M. D. & Petrie, M. Why do females mate multiply? A review of the genetic benefits. *Biol. Rev.* **75**, 21–64 (2000).
- Duellman, W. E. & Trueb, L. *Biology of Amphibians* 1–696 (The Johns Hopkins Univ. Press, Baltimore and London, 1986).
- Grossenbacher, K., *et al.* in *Atlas of Amphibians and Reptiles in Europe* (ed. Gasc, J. P.) 158–159 (Societas Europaea Herpetologica and Muséum National d'Histoire Naturelle (IEGB/SPN), Paris, 1997).
- Elmberg, J. Long-term survival, length of breeding season, and operational sex ratio in a boreal population of common frogs, *Rana temporaria* L. *Can. J. Zool.* **68**, 121–127 (1990).
- Byrne, P. G., Roberts, J. D. & Simmons, L. W. Sperm competition selects for increased testes mass in Australian frogs. *J. Evol. Biol.* **15**, 347–355 (2001).
- Gibbons, M. M. & McCarthy, T. K. The reproductive output of frogs *Rana temporaria* (L.) with particular reference to body size and age. *J. Zool.* **209**, 579–583 (1986).
- Elmberg, J. Factors affecting male yearly mating success in the common frog, *Rana temporaria*. *Behav. Ecol. Sociobiol.* **28**, 125–131 (1991).
- Elmberg, J. Random mating in a boreal population of European common frogs *Rana temporaria* L. *Holarc. Ecol.* **10**, 193–195 (1987).
- Foerster, K., Delhey, K., Johnsen, A., Liffield, J. T. & Kempaers, B. Females increase offspring

- heterozygosity and fitness through extra-pair matings. *Nature* **425**, 714–717 (2003).
26. Rico, C., Kuhnlein, U. & Fitzgerald, G. J. Male reproductive tactics in the threespine stickleback - an evaluation by DNA fingerprinting. *Mol. Ecol.* **1**, 79–87 (1992).
 27. Jones, A. G., Östlund-Nilsson, S. & Avise, J. C. A microsatellite assessment of sneaked fertilizations and egg thievery in the fifteen-spined stickleback. *Evolution* **52**, 848–858 (1998).
 28. Vos, C. C., De Jong, A. G., Goedhart, P. W. & Smulders, M. J. M. Genetic similarity as a measure for connectivity between fragmented populations of the moor frog (*Rana arvalis*). *Heredity* **86**, 598–608 (2001).
 29. Marshall, T. C., Slate, J., Kruuk, L. & Pemberton, J. M. Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* **7**, 639–655 (1998).
 30. Neff, B. D. & Pitcher, T. E. Assessing the statistical power of genetic analyses to detect multiple mating in fishes. *J. Fish Biol.* **61**, 739–750 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank T. J. C. Beebe, E. Hespeler and M. Fuerst for help with microsatellite techniques. The team of the Resposumo Refuge (Ursi, Dolores, David, Juani and Javier Abajo), X. González, J. Palanca and N. Palanca provided help and facilities. The Dirección General del Medio Natural de la Diputación General de Aragón issued research permits for the Circo de Piedrafita area. D.R.V. was supported by grants of the University of Vigo, Xunta de Galicia and Deutscher Akademischer Austauschdienst. Grants of the Deutsche Forschungsgemeinschaft to M.V. and A.M. and of the Instituto de Estudios Altoaragoneses to M.V. supported the laboratory work.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.R.V. (vieites@berkeley.edu).

Genomic analysis of regulatory network dynamics reveals large topological changes

Nicholas M. Luscombe^{1*}, M. Madan Babu^{4*}, Haiyuan Yu¹, Michael Snyder², Sarah A. Teichmann⁴ & Mark Gerstein^{1,3}

¹Department of Molecular Biophysics and Biochemistry, ²Department of Molecular, Cellular and Developmental Biology, ³Department of Computer Science, Yale University, PO Box 208114, New Haven, Connecticut 06520-8114, USA

⁴Division of Structural Studies, MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

* These authors contributed equally to this work

Network analysis has been applied widely, providing a unifying language to describe disparate systems ranging from social interactions to power grids. It has recently been used in molecular biology, but so far the resulting networks have only been analysed statically^{1–8}. Here we present the dynamics of a biological network on a genomic scale, by integrating transcriptional regulatory information^{9–11} and gene-expression data^{12–16} for multiple conditions in *Saccharomyces cerevisiae*. We develop an approach for the statistical analysis of network dynamics, called SANDY, combining well-known global topological measures, local motifs and newly derived statistics. We uncover large changes in underlying network architecture that are unexpected given current viewpoints and random simulations. In response to diverse stimuli, transcription factors alter their interactions to varying degrees, thereby rewiring the network. A few transcription factors serve as permanent hubs, but most act transiently only during certain conditions. By studying sub-network structures, we show that environmental responses facilitate fast signal propagation (for example, with short regulatory cascades), whereas the cell cycle and sporulation direct temporal progression through multiple stages (for example, with highly inter-connected transcription factors). Indeed, to drive the latter

processes forward, phase-specific transcription factors inter-regulate serially, and ubiquitously active transcription factors layer above them in a two-tiered hierarchy. We anticipate that many of the concepts presented here—particularly the large-scale topological changes and hub transience—will apply to other biological networks, including complex sub-systems in higher eukaryotes.

We began by assembling a static representation of known regulatory interactions from the results of genetic, biochemical and ChIP (chromatin immunoprecipitation)–chip experiments. Figure 1 illustrates the complexity of the resultant network, which contains 7,074 regulatory interactions between 142 transcription factors and 3,420 target genes (interactions can be between transcription factors and non-transcription factor targets, or two transcription factors). To get a dynamic perspective, we integrated gene-expression data for the following five conditions: cell cycle¹³, sporulation¹⁴, diauxic shift¹², DNA damage¹⁶ and stress response¹⁵. From these data, we traced paths in the regulatory network that are active in each condition using a trace-back algorithm (see Methods).

Figure 1b presents the sub-networks active under different cellular conditions, and gross changes are apparent in the distinct sections of the network that are highlighted. Recent functional genomics studies have analysed the dynamics of a few transcription factors^{17,18}; however, Fig. 1 represents the first dynamic view of a genome-scale network.

Half of the targets are uniquely expressed in only one condition; in contrast, most transcription factors are used across multiple processes. The active sub-networks maintain or rewire regulatory interactions, and over half of the active interactions (1,476 of 2,476 total) are completely supplanted by new ones between conditions. Only 66 interactions are retained across four or more conditions; these comprise ‘hot links’⁶ that are always on (compared with the rest of the network) and mostly regulate house-keeping functions.

The large number of changing interactions makes rigorous comparison of active sub-networks impossible visually. Consequently, we introduce SANDY, an approach that combines: standard measures of network connectivity (involving global topological statistics⁶ and local network motifs⁴), newly derived follow-on statistics, and comparisons against simulated controls to assess the significance of each observation.

Overall, our calculations divide the five condition-specific sub-networks into two categories: endogenous and exogenous (Fig. 1). This allows us to rationalize the different sub-network structures in terms of the biological requirements of each condition. Endogenous processes (cell cycle and sporulation) are multi-stage and operate with an internal transcriptional programme, whereas exogenous states (diauxic shift, DNA damage and stress response) constitute binary events that react to external stimuli with a rapid turnover of expressed genes.

We begin SANDY by examining global topological measures that quantify network architecture (Fig. 1c)⁶. The view from recent studies is that these statistics are remarkably constant across many biological networks (including regulatory systems)^{1,5,6,19,20}. Moreover, most of them remain invariant between randomly simulated sub-graphs of different sizes (Methods).

In fact, we show that topological measures change considerably between the endogenous and exogenous sub-networks. (The probability, P , that their topological measures originate from the same population is $<10^{-4}$; Supplementary Information.) Furthermore, most of the observed measurements differ significantly from random expectation and are insensitive to addition of noise in the underlying network (Methods). The ‘in-degree’ (k_{in}) is the number of incoming edges per node (that is, the number of transcription factors regulating a target). Its average across each sub-network decreases by 20% from endogenous to exogenous conditions ($P < 5 \times 10^{-3}$). The ‘out-degree’ (k_{out}) represents the number of

outgoing edges per node (that is, the number of target genes for each transcription factor). Average values double from endogenous to exogenous conditions ($P < 0.338$; Supplementary Information). The 'path length' (l) is the shortest distance between two nodes (here, it is the number of intermediate regulators between a transcription factor and a terminating target gene). Its average halves from endogenous to exogenous conditions ($P < 10^{-3}$). Finally, the 'clustering coefficient' (c) gauges the level of inter-connectivity around a node (that is, the level of transcription factor inter-regulation). Values range from 0 for totally dispersed nodes to 1 for fully connected ones. Average coefficients nearly halve from endogenous to exogenous conditions ($P < 10^{-3}$).

In biological terms, the small in-degrees for target genes in exogenous conditions indicate that transcription factors are regulating in simpler combinations, and the large out-degrees signify that each transcription factor has greater regulatory influence by targeting more genes simultaneously. The short paths imply faster propagation of the regulatory signal. Conversely, long paths in the multi-stage, endogenous conditions suggest slower action arising from the formation of regulatory chains to control intermediate phases. Finally, high clustering coefficients in endogenous conditions signify greater inter-regulation between transcription factors. In summary, sub-networks have evolved to produce rapid, large-scale responses in exogenous states, and carefully coordinated processes in endogenous conditions (Fig. 1a).

SANDY also examines the sub-networks locally by calculating the occurrence of network motifs⁴, which are compact, specific patterns of inter-connection between transcription factors and targets. We show the occurrence of the most common motifs in Fig. 1c: single-input, multiple-input, and feed-forward-loop motifs (SIMs, MIMs, and FFLs). In SIMs a single transcription factor targets many genes;

in MIMs multiple transcription factors co-regulate sets of genes; and in FFLs a primary transcription factor regulates a secondary one, and both target a final gene. Motifs appear at similar relative frequencies across regulatory networks of diverse organisms (although individual motifs are not conserved)⁷, and this is also true for the randomly simulated sub-graphs. Therefore, constancy in motif usage is expected across conditions.

However, Fig. 1c shows that the relative occurrence of motifs varies considerably between endogenous and exogenous conditions ($P < 10^{-3}$). SIMs are favoured in exogenous sub-networks where they comprise >55% of regulatory interactions in motifs. But the frequency drops to ~35% in endogenous processes. Instead, these states favour FFLs (~44%). MIMs do not significantly change in their frequency.

Previous studies defined precise regulatory properties and information processing tasks for motifs⁴. SIMs and MIMs are implicated in conferring similar regulation over groups of genes, so they are ideal for directing the large-scale gene activation found in exogenous conditions. FFLs are buffers that respond only to persistent input signals. They are suited for endogenous conditions, as cells cannot initiate a new stage until the previous one has stabilized. Though used sparingly in exogenous processes they may be important in filtering spurious external stimuli.

Having quantified global and local changes with standard topological measures, we then moved to the follow-on statistics in SANDY (Fig. 2). Like many large-scale networks, the regulatory system displays scale-free characteristics (the probability P_k that a transcription factor targets k genes is proportional to $k^{-\gamma}$ for constant γ). This behaviour (maintained across all active sub-networks) signifies the presence of regulatory hubs targeting disproportionately large numbers of genes. Hubs are of general interest

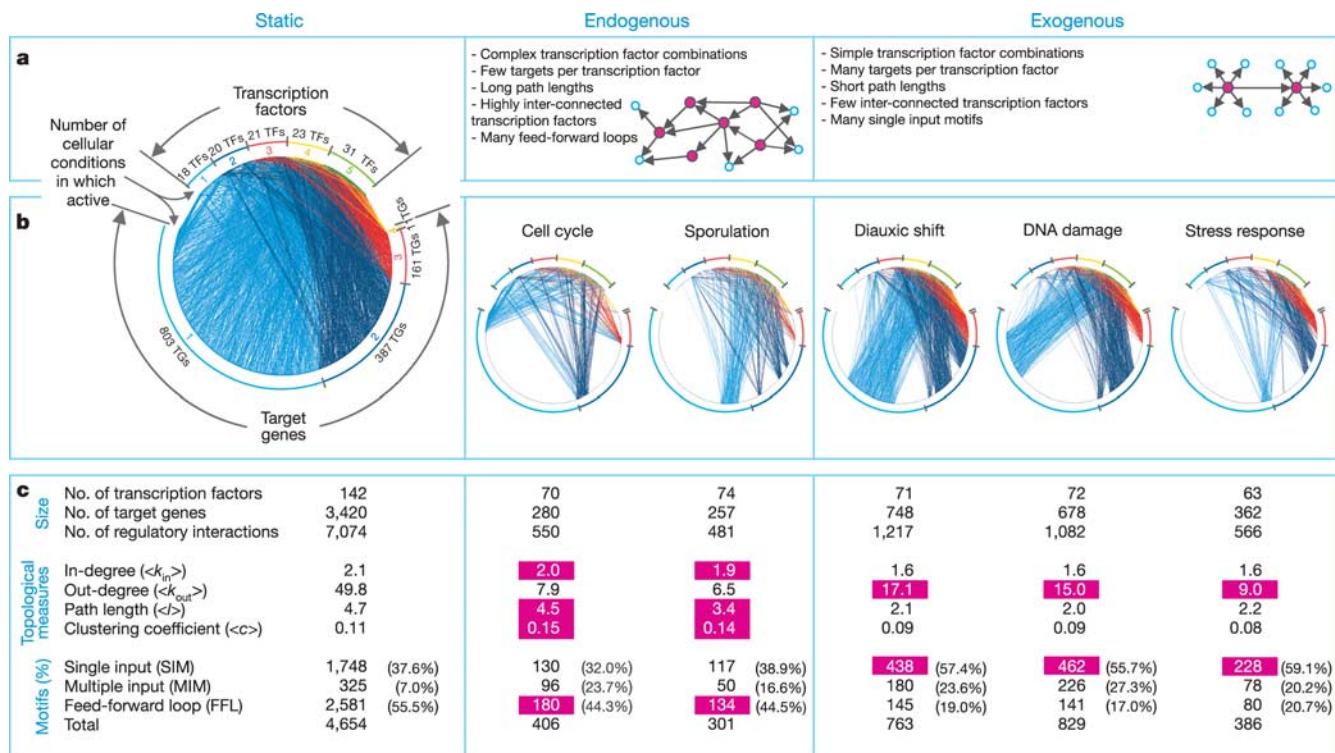


Figure 1 Dynamic representation of the transcriptional regulatory network and standard statistics. **a**, Schematics and summary of properties for the endogenous and exogenous sub-networks. **b**, Graphs of the static and condition-specific networks. Transcription factors and target genes are shown as nodes in the upper and lower sections of each graph respectively, and regulatory interactions are drawn as edges; they are coloured by the number of conditions in which they are active. Different conditions use distinct

sections of the network. **c**, Standard statistics (global topological measures and local network motifs) describing network structures. These vary between endogenous and exogenous conditions; those that are high compared with other conditions are shaded. (Note, the graph for the static state displays only sections that are active in at least one condition, but the table provides statistics for the entire network including inactive regions.)

as they represent the most influential components of a network⁶ and, accordingly, tend to be essential²¹. They are thought to target a broad spectrum of gene functions^{4,11,22} and are commonly located upstream in the network² to expand their influence via secondary transcription factors²³. These observations suggest that hubs would be invariant features of the network across conditions, and this expectation is supported by the random simulations that converge on similar sets of transcription factor hubs.

Figure 2a shows the observed regulatory hubs in each of the five conditions (Methods). They divide into two groups. The smaller one represents permanent hubs, which, in line with expectation, are important regardless of cellular state. They mainly comprise multi-functional transcription factors (such as Abf1) and house-keeping regulators (such as Mig1/2), and are responsible for maintaining the

hot links. However, contrary to expectation, most hubs (78%) are transient; that is, they are influential in one condition, but less so in others. Exogenous conditions have fewer transient hubs, suggesting a more centralized command structure. (This is reflected in different γ ; Supplementary Information.) About half of these hubs are known to be important for their respective conditions (for example, Swi4 in the cell cycle; Methods). For the remainder with sparse annotations, their transient-hub status in a particular condition considerably augments their functional annotation (for example, Sok2 in the cell cycle). These hubs may also relate to condition-dependent lethality, and this has clear implications for identifying specific drug targets.

The defining feature of transient hubs is their capacity to change interactions between conditions. We attempted to quantify this

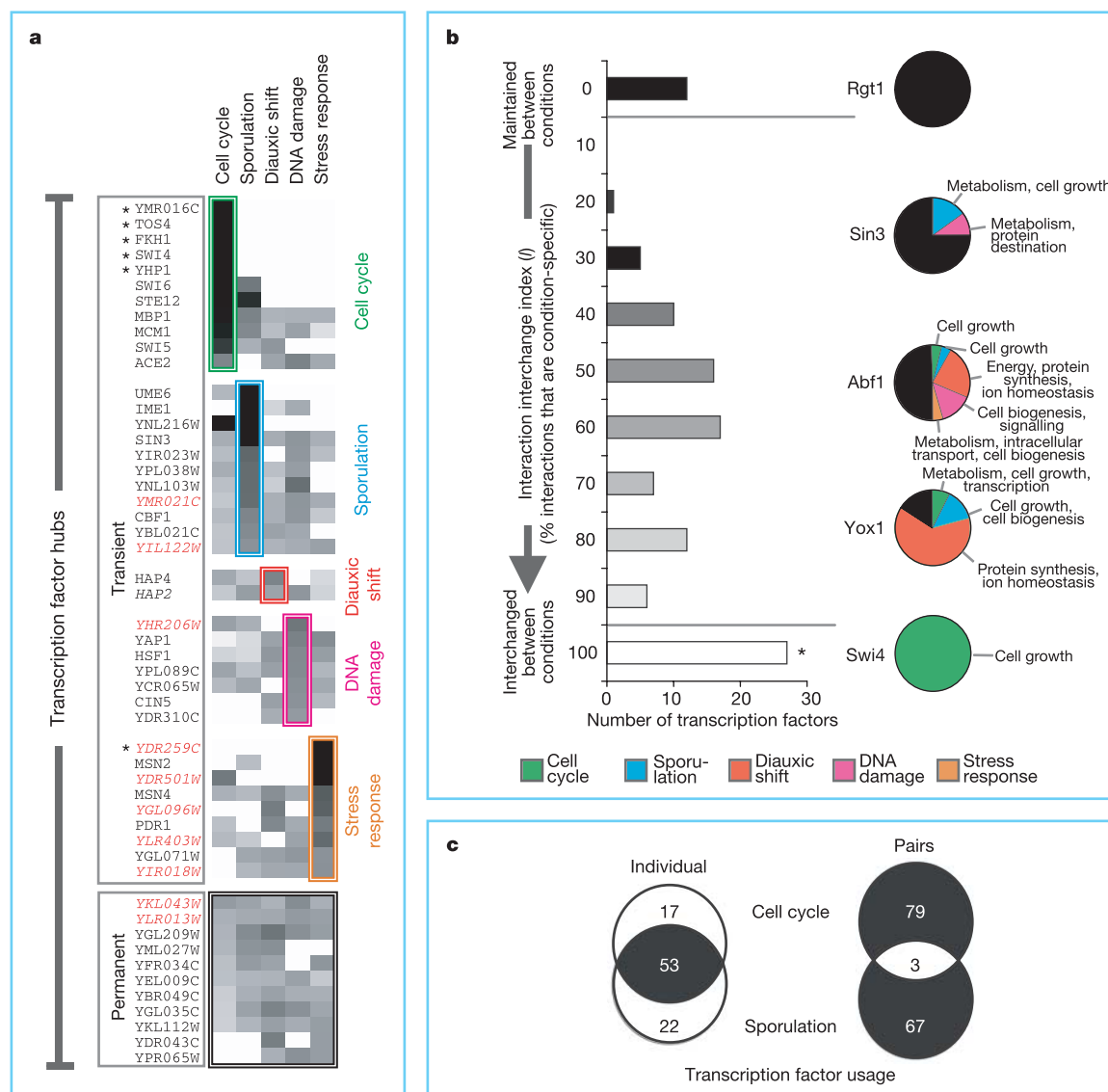


Figure 2 Newly derived 'follow-on' statistics for network structures. **a**, Transcription factor hub usage in different cellular conditions. The cluster diagram shades cells by the normalized number of genes targeted by transcription factor hubs in each condition. One cluster represents permanent hubs and the others condition-specific transient hubs. Genes are labelled with four-letter names when they have an obvious functional role in the condition, and seven-letter open reading frame names when there is no obvious role. Of the latter, gene names are red and italicised when functions are poorly characterized. Starred hubs show extreme interchange index values, $I = 1$. **b**, Interaction interchange (I) of transcription factors between conditions. A histogram of I for all active transcription

factors shows a uni-modal distribution with two extremes. Pie charts show five example transcription factors with different proportions of interchanged interactions. We list the main functions of the distinct target genes regulated by each example transcription factor. Note how the transcription factors' regulatory functions change between conditions. **c**, Overlap in transcription factor usage between conditions. Venn diagrams show the numbers of individual transcription factors (large intersection) and pair-wise transcription factor combinations (small intersection) that overlap between the two endogenous conditions.

rewiring more broadly for every transcription factor in the network with the interchange index, I . This is defined so that higher values associate with transcription factors replacing a larger fraction of their interactions. Its histogram reveals a uni-modal central distribution with two groups of extreme outliers (Fig. 2b). At one extreme ($I \leq 10\%$), 12 transcription factors retain all interactions across multiple states. At the other end ($I \geq 90\%$), 27 transcription factors replace all interactions in switching conditions. Many of these are so extreme that they only regulate genes in a single condition and are inactive otherwise. These include six transient hubs of known importance for the cell cycle and stress response. Most transcription factors interchange only part of their interactions ($10\% < I < 90\%$). This group comprises most of the hubs; surprisingly, permanent hubs interchange interactions as often as transient ones, but over a larger number of conditions. Furthermore, transcription factors in this group often regulate genes of distinct functions in different conditions, thus shifting regulatory roles. For example, the permanent hub Abf1 regulates cell growth during endogenous conditions, but refocuses to intracellular transport during stress response (in addition to its maintained core functions).

The rewiring highlighted by the interchange index allows transcription factors to be active in many conditions. Indeed, 95 of the 142 transcription factors are used in more than one process (Fig. 1b). Specifically, within endogenous conditions 53 of 92 transcription

factors overlap between cell cycle and sporulation (Fig. 2c), and there is a similar overlap for exogenous conditions (Supplementary Information). With so much intersection in the repertoire of active transcription factors, the precise regulation of a condition cannot arise from the specificity of individual transcription factors. As others have observed²⁴, combinatorial transcription factor usage seems to be the key. We calculated that there are 360 unique pairwise transcription factor combinations (that is, two transcription factors regulating the same target) used in at least one condition. In contrast to individual transcription factors, only a minor proportion of pairs (51 of 360) participate in multiple processes and just 3 of 149 pairs overlap between endogenous conditions (Fig. 2c).

Thus far we have focused on the large dynamic changes occurring between different cellular conditions. However, dynamic transitions also take place within individual processes. Earlier, SANDY defined endogenous sub-networks by their long paths and high clustering. We can study the source of these observations by looking at the full scope of inter-regulation between transcription factors during the cell cycle (Fig. 3). This is possible because in a previous study¹³ expression-level measurements were made throughout the cell cycle and differentially expressed genes were assigned to one of five phases (early G1, late G1, S, G2 and M). We then traced back from the classified genes to identify active sub-networks during each phase (Methods).

A cluster diagram (Fig. 3a) shows that most transcription factors that are active in the cell cycle operate only in a particular phase (for example, Swi4 in late G1). Additionally, a sizeable minority of transcription factors is ubiquitously active throughout the whole cycle. We uncover two major forms of transcription factor inter-regulation. In serial inter-regulation²⁵ (Fig. 3b), the phase-specific transcription factors regulate each other in a sequential manner to drive the cell cycle forward. In fact, we detect complete loops of interactions within the complex circuitry, and the resulting regulatory cascades undoubtedly create the long paths. We also introduce the concept of parallel inter-regulation (Fig. 3c), in which the ubiquitous transcription factors control the phase-specific ones in a two-tiered system. This effectively provides a stable signal to aid the transition between phases. Furthermore, because about a third of ubiquitous transcription factors comprise permanent hubs, they may provide a channel of communication to relate the cell-cycle progression with house-keeping functions. Similar observations apply to sporulation (Supplementary Information).

SANDY presents an approach to examine biological network dynamics. In applying it to the yeast regulatory system, it becomes apparent that many observations made in the static state are not applicable to a dynamic perspective, we uncover substantial topological changes in network structure, and we capture the essence of the transcriptional regulatory data in a new way. Because of limitations in current data sets, we can examine this only through integrating gene-expression information. However, we anticipate future experiments to determine condition-specific interactions directly. Given the robustness of the observations to large perturbations (Methods), we expect our approach and findings to remain valid for these new data sets. Furthermore, we anticipate that many of the concepts we have introduced could be readily transferred to other types of biological networks and complex sub-systems in multicellular organisms, such as those directing the circadian cycle²⁶ and cellular development. □

Methods

Detailed descriptions of the methods are in the Supplementary Information and at <http://sandy.topnet.gersteinlab.org>.

Data sets

The transcriptional regulatory network was assembled from the results of genetic, biochemical and ChIP-chip experiments^{9–11}. The gene-expression data were compiled from 240 microarray experiments for five conditions^{12–16}. We identified the following

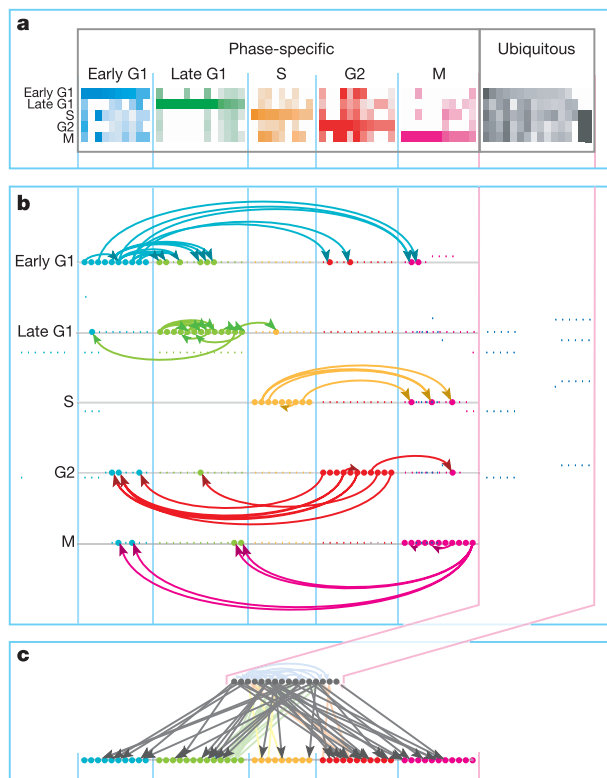


Figure 3 Transcription factor inter-regulation during the cell cycle time-course. **a**, The 70 transcription factors active in the cell cycle. The diagram shades each cell by the normalized number of genes targeted by each transcription factor in a phase. Five clusters represent phase-specific transcription factors and one cluster is for ubiquitously active transcription factors. Note, both hub and non-hub transcription factors are included. Transcription factor names are given in the Supplementary Information. **b**, Serial inter-regulation between phase-specific transcription factors. Network diagrams show transcription factors that are active in one phase regulate transcription factors in subsequent phases. In the late phases, transcription factors apparently regulate those in the next cycle. **c**, Parallel inter-regulation between phase-specific and ubiquitous transcription factors in a two-tiered hierarchy. Serial and parallel inter-regulation operate in tandem to drive the cell cycle while balancing it with basic house-keeping processes.

numbers of genes with differential expression: cell cycle, 455; sporulation, 477; diauxic shift, 1,823; DNA damage, 1,718; and stress response, 866.

Trace-back algorithm

We used the following algorithm to define sections of the regulatory network used in each condition: (1) identify transcription factors as being 'present' in a condition if they have sufficiently high expression levels; (2) flag differentially expressed genes that appear in the regulatory network; (3) mark as 'active' the regulatory links between present transcription factors and differentially expressed genes; and (4) search for any other present transcription factors that are linked to a transcription factor with an already active link and make this connection active. The last step is repeated until no more links are made active. The same procedure identifies sub-networks that are active in particular phases of cell cycle and sporulation.

SANDY

This extends the methodology used by the TopNet software tool²⁷ and it evaluates each sub-network with the following: (1) standard statistics, including global measures of topology (k_{in} , k_{out} , I , and c^6) and local motif occurrence (SIM, MIM and FFL)⁴; (2) follow-on statistics, including permanent and transient hub identification, interchange index (I) and counting the overlap in transcription factor usage (individual and pairs) across multiple conditions. (Hubs are transcription factors in the top 30%, by number of target genes, in at least one condition. The number of target genes is normalized to measure the relative influence of a transcription factor hub in a particular process.) In all cases, regulatory functions are obtained from the *Saccharomyces Genome Database*²⁸ and are current as of June 2004; and (3) a comparison of observations and random expectation by simulating sub-graphs that are similar in size to each sub-network, and calculating standard and follow-on statistics for them. Simulated sub-graphs sample the same number of differentially expressed genes and trace through the static network. We also tested the sensitivity of our observations to noise by randomly perturbing the static networks by 30% (random addition, deletion and replacement of interactions), tracing back from the original differentially expressed genes and then recalculating the statistics.

Received 15 January; accepted 24 June 2004; doi:10.1038/nature02782.

- Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A. L. The large-scale organization of metabolic networks. *Nature* **407**, 651–654 (2000).
- Guelzim, N., Bottani, S., Bourguin, P. & Kepes, F. Topological and causal structure of the yeast transcriptional regulatory network. *Nature Genet.* **31**, 60–63 (2002).
- Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
- Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genet.* **31**, 64–68 (2002).
- Oltvai, Z. N. & Barabási, A. L. Systems biology: Life's complexity pyramid. *Science* **298**, 763–764 (2002).
- Barabási, A. L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nature Rev. Genet.* **5**, 101–113 (2004).
- Milo, R. *et al.* Superfamilies of evolved and designed networks. *Science* **303**, 1538–1542 (2004).
- Teichmann, S. A. & Babu, M. M. Gene regulatory network growth by duplication. *Nature Genet.* **36**, 492–496 (2004).
- Svetlov, V. V. & Cooper, T. G. Review: compilation and characteristics of dedicated transcription factors in *Saccharomyces cerevisiae*. *Yeast* **11**, 1439–1484 (1995).
- Horak, C. E. *et al.* Complex transcriptional circuitry at the G1/S transition in *Saccharomyces cerevisiae*. *Genes Dev.* **16**, 3017–3033 (2002).
- Lee, T. I. *et al.* Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799–804 (2002).
- DeRisi, J. L., Iyer, V. R. & Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686 (1997).
- Cho, R. J. *et al.* A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell* **2**, 65–73 (1998).
- Chu, S. *et al.* The transcriptional program of sporulation in budding yeast. *Science* **282**, 699–705 (1998).
- Gasch, A. P. *et al.* Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**, 4241–4257 (2000).
- Gasch, A. P. *et al.* Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p. *Mol. Biol. Cell* **12**, 2987–3003 (2001).
- Odom, D. T. *et al.* Control of pancreas and liver gene expression by HNF transcription factors. *Science* **303**, 1378–1381 (2004).
- Zeitlinger, J. *et al.* Program-specific distribution of a transcription factor dependent on partner transcription factor and MAPK signaling. *Cell* **113**, 395–404 (2003).
- Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
- Wagner, A. & Fell, D. A. The small world inside large metabolic networks. *Proc. R. Soc. Lond. B* **268**, 1803–1810 (2001).
- Yu, H., Greenbaum, D., Xin Lu, H., Zhu, X. & Gerstein, M. Genomic analysis of essentiality within protein networks. *Trends Genet.* **20**, 227–231 (2004).
- Martínez-António, A. & Collado-Vides, J. Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr. Opin. Microbiol.* **6**, 82–89 (2003).
- Madan Babu, M. & Teichmann, S. A. Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nucleic Acids Res.* **31**, 1234–1244 (2003).
- Pilpel, Y., Sudarsanam, P. & Church, G. M. Identifying regulatory networks by combinatorial analysis of promoter elements. *Nature Genet.* **29**, 153–159 (2001).
- Simon, I. *et al.* Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell* **106**, 697–708 (2001).
- Ueda, H. R. *et al.* A transcription factor response element for gene expression during circadian night. *Nature* **418**, 534–539 (2002).

- Yu, H., Zhu, X., Greenbaum, D., Karro, J. & Gerstein, M. TopNet: a tool for comparing biological sub-networks, correlating protein properties with topological statistics. *Nucleic Acids Res.* **32**, 328–337 (2004).
- Christie, K. R. *et al.* *Saccharomyces Genome Database (SGD)* provides tools to identify and analyze sequences from *Saccharomyces cerevisiae* and related sequences from other organisms. *Nucleic Acids Res.* **32**, D311–D314 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank P. Bertone, N. Domedel-Puig, E. Hovig, R. Jansen, K. Kleivi, G. Koentges, E. Koonin, B. Lenhard, A. Paccanaro, J. Rozowsky, J. Tegner, V. Trifonov, A. Todd, Y. Xia and H. Zao for comments on the paper. N.M.L. thanks the Anna Fuller Fund and the MRC LMB Visitor's Program. M.M.B. acknowledges financial support from the Cambridge Commonwealth Trust, Trinity College, Cambridge and the MRC LMB. M.G. is supported by the NSF and NIH.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.M.L., M.M.B., S.A.T. & M.G. (sandy@bioinfo.mbb.yale.edu).

Long-lasting self-inhibition of neocortical interneurons mediated by endocannabinoids

Alberto Bacci, John R. Huguenard & David A. Prince

Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, California 94305, USA

Neocortical GABA-containing interneurons form complex functional networks responsible for feedforward and feedback inhibition and for the generation of cortical oscillations associated with several behavioural functions^{1,2}. We previously reported that fast-spiking (FS), but not low-threshold-spiking (LTS), neocortical interneurons from rats generate a fast and precise self-inhibition mediated by inhibitory autaptic transmission³. Here we show that LTS cells possess a different form of self-inhibition. LTS, but not FS, interneurons undergo a prominent hyperpolarization mediated by an increased K^+ -channel conductance. This self-induced inhibition lasts for many minutes, is dependent on an increase in intracellular $[Ca^{2+}]$ and is blocked by the cannabinoid receptor antagonist AM251, indicating that it is mediated by the autocrine release of endogenous cannabinoids. Endocannabinoid-mediated slow self-inhibition represents a powerful and long-lasting mechanism that alters the intrinsic excitability of LTS neurons, which selectively target the major site of excitatory connections onto pyramidal neurons; that is, their dendrites^{4–7}. Thus, modulation of LTS networks after their sustained firing will lead to long-lasting changes of glutamate-mediated synaptic strength in pyramidal neurons, with consequences during normal and pathophysiological cortical network activities.

We obtained whole-cell recordings from fast-spiking and LTS interneurons in layer V of rat somatosensory neocortical slices. Cells were characterized electrophysiologically by their firing behaviours as described previously⁸ (Fig. 1a, c, insets; see also Methods). All biocytin-filled LTS interneurons contained cholecystokinin (CCK) ($n = 7$; Supplementary Fig. S1), whereas FS cells were CCK-negative ($n = 4$; Supplementary Fig. S1), indicating that we sampled a homogeneous non-FS cell population.

Although LTS interneurons lack GABAergic autaptic transmission³, we found a unique form of slow self-inhibition (SSI) in

42 of 43 LTS cells. Multiple trains of evoked action potentials (APs) (first and last train in series shown in Fig. 1b, d) were followed by a prominent hyperpolarization of resting membrane potential ($V_m = -60.5 \pm 0.6$ mV versus -69.2 ± 1.0 mV, before and 5 min after DC-induced firing; $n = 9$, $p \ll 0.0001$; Fig. 1a, e). The hyperpolarization was long-lasting (more than 35 min), occurred selectively in LTS cells (V_m in FS cells = -65.7 ± 1.9 mV versus -65.6 ± 2.0 mV, before and 5 min after DC-induced firing; $n = 6$, $p > 0.9$; Figs 1c, e and 2d), and occurred in the presence of synaptic blockers (Methods). In a few LTS cells there was a partial recovery of V_m after 12–15 min (not shown). SSI onset was well fitted by an exponential function with a time constant of 43 s.

LTS cells tend to fire at relatively low frequencies (20–50 Hz)

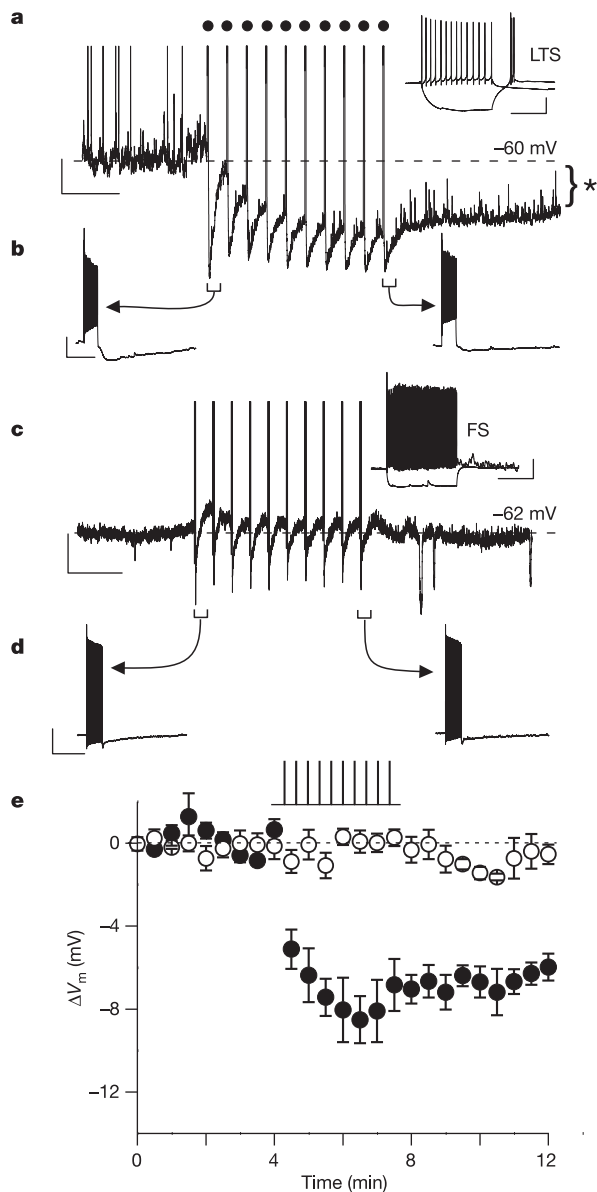


Figure 1 SSI occurs in LTS, not FS, neocortical interneurons. **a**, LTS interneuron responds to +600-pA current pulses (circles) with AP trains and long-lasting hyperpolarization (asterisk). Inset: LTS firing behaviour. **b**, First and last responses from **a**. **c**, **d**, As in **a** and **b**, but in a FS cell; SSI was not induced. Inset: FS firing behaviour. APs here and later have been truncated for display purposes. **d**, First and last responses from **c**. **e**, Changes in V_m (ΔV_m) in LTS cells (filled circles, $n = 9$) and FS cells (open circles, $n = 6$). Vertical lines indicate current pulses. Hyperpolarization was induced selectively in LTS cells. Scale bars, 1 min and 5 mV (**a**, **c**), 250 ms and 20 mV (insets), 2.5 s and 20 mV (**b**, **d**). Error bars indicate s.e.m.

during network oscillations⁹. To test the dependence of SSI on firing frequency within and outside this range, we applied brief suprathreshold depolarizing current steps to evoke trains of 10-Hz or 50-Hz APs (Fig. 2a, b). Both 10-Hz and 50-Hz spike trains evoked prominent SSI, and responses were comparable in amplitude and duration (Fig. 2c, g), indicating that even relatively low-frequency firing can trigger self-inhibition in LTS interneurons. Neither 10-Hz (Fig. 2d, g) nor 50-Hz stimuli (not shown) induced SSI in FS interneurons. SSI was accompanied by a marked, long-lasting increase in membrane conductance (g_m) (2.9 ± 0.5 nS in control versus 4.6 ± 0.5 and 5.1 ± 0.8 nS after 3 and 10 min of SSI induction, respectively; $n = 9$, $p < 0.002$ in both cases; Figs 2c, e, h). FS interneurons showed no significant changes in g_m from

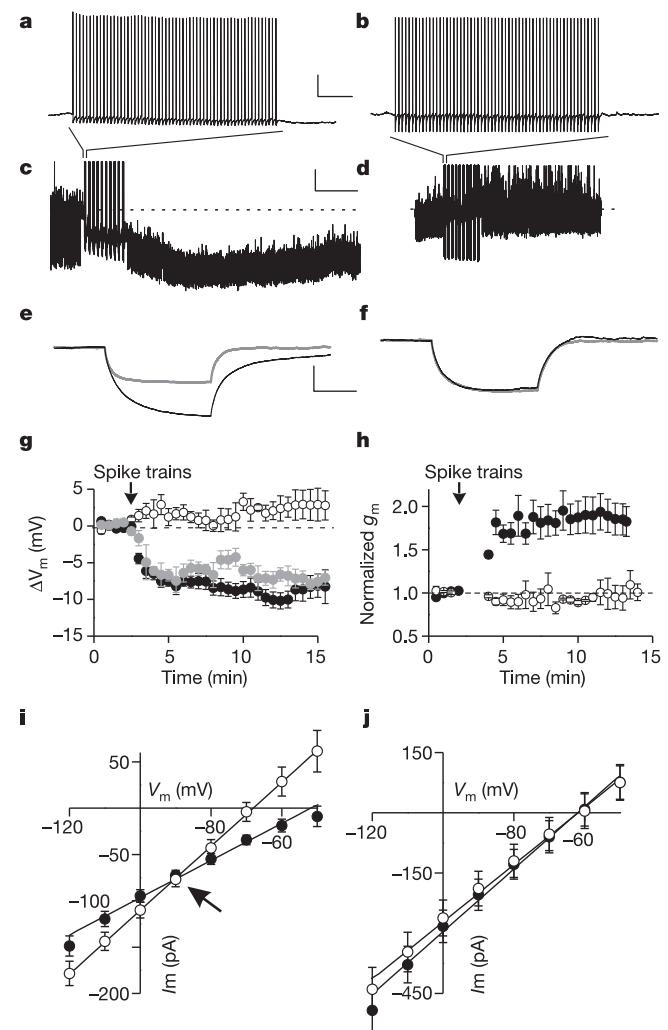


Figure 2 Low firing frequencies induce SSI, which reflects an increased membrane conductance. **a**, **b**, Brief depolarizing pulses reliably evoke LTS (**a**) and FS (**b**) spikes. **c**, **d**, Ten trains of 10-Hz APs elicited SSI in LTS (**c**) but not FS (**d**) cell. Negative deflections in **c**–**f** show conductance tests. **e**, **f**, Increased g_m (smaller responses) followed AP trains in LTS (**e**) but not FS (**f**) interneuron. Black traces, before trains; grey traces, after trains. **g**, Plot of ΔV_m against time during 50-Hz train stimulation (LTS (black filled circles), $p < 0.0001$, $n = 12$) and 10-Hz train stimulation (LTS (grey filled circles), $p < 0.01$, $n = 6$; FS (open circles), $p > 0.5$, $n = 4$). **h**, Normalized g_m time course in LTS (filled circles, $n = 9$) and FS (open circles, $n = 4$) interneurons. **i**, **j**, Current–voltage relationships in LTS (**i**, $n = 7$) and FS (**j**, $n = 6$) cells before (filled circles) and after (open circles) trains. The arrow indicates the intersection of current–voltage plots in LTS cells. Scale bars, 20 mV and 1 s (**a**, **b**), 5 mV and 3 min (**c**, **d**), 5 mV and 100 ms (**e**, **f**). Error bars indicate s.e.m.

control when exposed to equivalent stimuli (3.8 ± 0.7 nS in control versus 3.6 ± 0.6 and 3.8 ± 0.8 nS after 3 and 10 min of SSI induction, respectively; $n = 4$, $p > 0.7$ in both cases; Fig. 2d, f, h).

The conductance underlying SSI was examined by using voltage-clamp recordings of membrane currents at different holding potentials. Current–voltage plots before and after stimulus trains intersected at about -90 mV, near the calculated K^+ equilibrium potential (Fig. 2i), but far from E_{Cl^-} (about -16 mV), indicating activation of a K^+ conductance. Identical protocols applied to FS interneurons revealed no shifts in current–voltage relationships (Fig. 2j). Both the hyperpolarization and enhancement of g_m were reversed by the application of 0.1 mM Ba^{2+} after SSI induction (Supplementary Fig. S2), indicating that G-protein-coupled inwardly rectifying K^+ (GIRK) channels¹⁰ are activated during SSI.

Sustained neuronal firing results in elevations of intracellular $[Ca^{2+}]$ that might contribute to SSI either by directly activating a Ca^{2+} -dependent membrane conductance or by enhancing the release of a transmitter/modulator that acts on autoreceptors of the stimulated cell^{3,11}. We tested the role of increases in intracellular

$[Ca^{2+}]$ in LTS self-inhibition by including the fast Ca^{2+} chelator BAPTA (10 mM) in the whole-cell pipette and assessing effects on SSI evoked as above. Intracellular BAPTA prevented the induction of SSI in LTS cells (Fig. 3a, c; $n = 6$), as did extracellular application of Cd^{2+} (200 μ M) in the presence of control intracellular solution (Fig. 3b, c; $n = 6$), indicating that an elevation of $[Ca^{2+}]_i$ through voltage-dependent calcium channels triggers SSI in LTS cells. APs themselves were not required for SSI, because SSI could be induced in the presence of the Na^+ -channel blocker tetrodotoxin (Supplementary Fig. S4).

We proposed that SSI might be induced by a Ca^{2+} -dependent release of a diffusible factor, acting on the same cells in an autocrine fashion. A prominent candidate would be an endocannabinoid, because the neuronal isoform of the G-protein-coupled cannabinoid receptor CB1 is selectively expressed by CCK-containing interneurons¹², and the LTS cells we recorded express this peptide (Supplementary Fig. S1). In addition, the synthesis of endocannabinoids is highly dependent on intracellular calcium elevations^{12–14} that were presumably blocked along with SSI in our BAPTA and Cd^{2+} experiments. Moreover, activation of CB1 receptors leads to an increase of GIRK channel function^{12,15}.

To test the role of endocannabinoid receptors in LTS self-inhibition, we evoked spike trains as above, during local perfusion of the CB1 antagonist AM251 (ref. 16). AM251 (5 μ M) blocked SSI induction in LTS cells (Fig. 4a, b; $n = 8$) as well as stimulus-induced

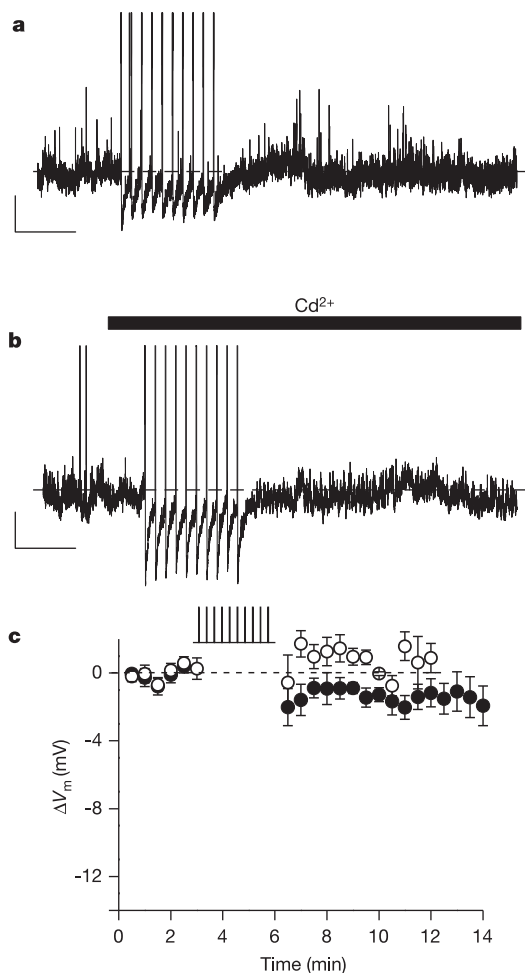


Figure 3 Intracellular increases in $[Ca^{2+}]$ through voltage-dependent Ca^{2+} channels are required for the induction of SSI in LTS cells. **a**, Response to SSI-inducing stimuli applied about 5 min into a whole-cell recording with intracellular BAPTA. **b**, Extracellular Cd^{2+} (200 μ M) prevented SSI induction during a whole-cell recording, in the presence of intracellular EGTA. Scale bars, 5 mV and 2 min. **c**, V_m time course in LTS cells exposed to SSI stimuli in the presence of intracellular BAPTA (filled circles) or extracellular Cd^{2+} (open circles). Each treatment completely blocked SSI in all tested cells ($p > 0.4$ and $p > 0.1$ in BAPTA and Cd^{2+} , respectively; $n = 6$ in both cases). Error bars indicate s.e.m.

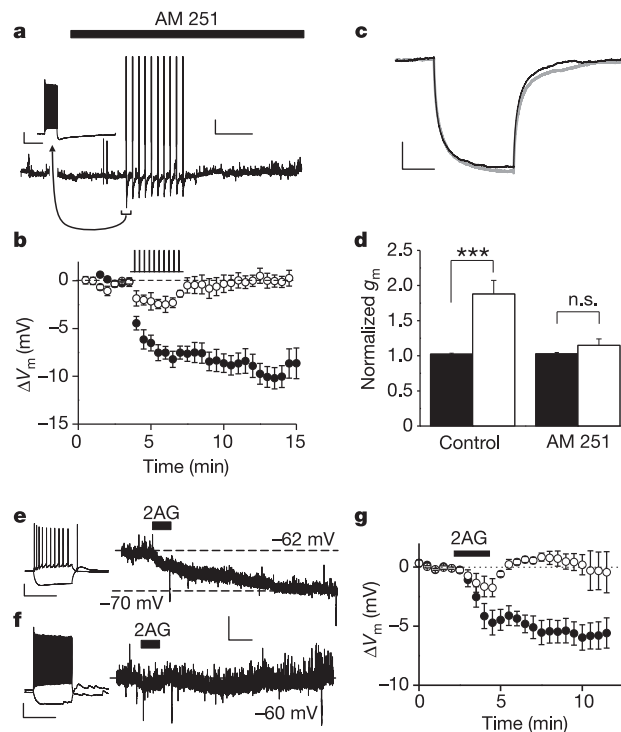


Figure 4 SSI is mediated by endocannabinoids. **a**, AM251 (5 μ M) prevented SSI. Inset: response to first depolarizing pulse. **b**, ΔV_m time course in control (filled circles, $n = 12$) and AM251-treated (open circles, $n = 8$) cells; AM251 prevented SSI in all cells ($p > 0.7$, $n = 8$). Control data are from Fig. 2g. **c**, Conductance tests before (black trace) and after (grey trace) SSI-inducing stimuli in AM251. **d**, Normalized g_m in control and AM251-treated cells, before (filled columns) and after (open columns) spike trains. The increase in g_m (asterisks indicate $p < 0.002$; $n = 9$) was prevented by AM251 ($p > 0.1$; $n = 5$). **e**, **f**, Left: Current-clamp identification of LTS (**e**) and FS (**f**) cells. Right: responses to applications of 2AG (10 μ M) in the same cells. **g**, Time course of the 2AG effect on V_m in LTS cells (filled circles, $p < 0.002$, $n = 7$) and FS cells (open circles, $p > 0.06$, $n = 6$). Scale bars, 5 mV and 2 min (**a**; **e** and **f** right), 20 mV and 500 ms (**a** inset; **e** and **f** left), 5 mV and 200 ms (**c**). Error bars indicate s.e.m.

increases in g_m (Fig. 4c, d). Moreover, AM251 applied after SSI induction partly reversed both the hyperpolarization and the change in g_m , indicating persistent CB1 receptor signalling during SSI expression (Supplementary Fig. S3). The relatively short (about 20-s) hyperpolarization that followed each evoked burst of APs in LTS and FS cells (Fig. 1a, c) was not related to SSI, because it was not altered by any of the above pharmacological manipulations (data not shown).

Finally, exogenous application of an endogenous cannabinoid, 2-arachidonylglycerol (2AG; 10 μ M)¹², resulted in a robust hyperpolarization similar in amplitude and time course to that induced by spike trains (Fig. 4e, g). The 2-AG-mediated hyperpolarization was not elicited in FS cells (Fig. 4f, g; $n = 6$), indicating that this subtype of interneuron does not express GIRK-coupled somatodendritic CB1 receptors. This finding does not exclude the presence of such receptors on axonal terminals¹².

Here we show a novel form of self-inhibition in CCK-containing LTS cortical interneurons. SSI does not depend on glutamatergic and/or GABAergic neurotransmission, but rather on an activity-dependent long-lasting hyperpolarization due to the activation of a GIRK conductance, with the resultant depression of LTS interneuron excitability (see Supplementary Table 1 and Supplementary Fig. S3d). This plastic change in intrinsic LTS cell excitability is triggered by a Ca^{2+} -dependent autocrine action of endocannabinoids, which are probably synthesized by and targeted to the same neurons.

Recent studies have shown that endocannabinoids released by postsynaptic neurons mediate retrograde signalling at hippocampal and cerebellar synapses, underlying the transient (15–20 s) phenomena of depolarization-induced suppression of inhibition (DSI) and excitation (DSE)^{16–18}. The proposed mechanism involves a negative modulation of presynaptic Ca^{2+} channels and transmitter release^{12,16–18}. In hippocampus and neocortical layers II and III, DSI is due to retrograde effects of endocannabinoids on perisomatic terminals of basket cells containing CCK^{2,19,20}.

Our results differ from those involving DSI and DSE in several important respects. In neocortical layer V, repetitive firing of LTS interneurons at frequencies within their physiological range⁹ induces a previously unknown endocannabinoid-mediated autocrine response—a persistent decrease in intrinsic excitability in the same LTS cells—that does not require retrograde signalling. In this model the highly lipophilic endocannabinoids, once synthesized, might affect nearby CB1 receptors by means of a localized intramembrane signalling mechanism²¹. The large Ba^{2+} -sensitive increases in g_m and the reversal potential of the underlying current indicate that, in contrast to the situation in DSI, CB1 receptors are present on neuronal somatodendritic membrane and are functionally coupled to K^+ channels. Recent data indicate that presynaptic K^+ channels might have a function in cerebellar DSI²², and retrograde (but not autocrine) cannabinoid signalling also affects K^+ channels in cerebellar interneurons²³, although with kinetics that are shorter than SSI.

A long-lasting self-induced hyperpolarization similar to SSI might underlie two recently reported endocannabinoid-mediated effects in the hippocampus, namely a long-term depression of GABAergic transmission after tetanic stimulation²⁴, and persistent endocannabinoid signalling affecting the short-term plasticity of presynaptic GABA release from CB1-expressing interneurons¹⁹.

Mechanisms underlying the long duration of SSI might include a sustained activation of CB1 receptors, as demonstrated by partial reversal of SSI with late AM251 applications (see Supplementary Fig. S3). Persistent endocannabinoid signalling could result from either altered metabolism through the catalytic enzyme fatty acid amide hydrolase, which is known to terminate endocannabinoid signalling^{12,25}, or to a relatively persistent change in CB1 activation state. A long-term downstream modification of the activated K^+ conductance, as might result from an alteration in channel phos-

phorylation²⁶, could contribute to SSI and might explain the incomplete effects of late AM251 application (Supplementary Fig. S3).

LTS and FS interneurons are connected through both electrical junctions and GABAergic synapses^{27,28}, forming two neocortical networks^{9,28,29} with crucial roles in setting and maintaining cortical oscillations². LTS interneuron axons target pyramidal cell dendrites^{4–6}, where most excitatory synapses are located⁷. The selective loss of LTS cell-evoked pyramidal dendritic inhibition during repetitive interneuronal discharge might therefore result in a prolonged enhancement of glutamatergic excitation onto cortical principal cells, altering their input–output relations and their participation in cortical network activities. Endocannabinoid-mediated, activity-dependent hyperpolarization in LTS cells might also have a crucial function in pyramidal cell disinhibition during epileptiform or other intense network activities known to be associated with the generation of high-frequency pyramidal cell and interneuronal firing^{9,30}. Prolonged plastic changes in cortical signal processing mediated by SSI of LTS cells could be one important mechanism through which activation of CB1 receptors by the recreational drug marijuana gives rise to its psychotropic effects¹². □

Methods

Preparation and electrophysiology of brain slices *in vitro*

Sprague–Dawley rats aged between postnatal day 13 (P13) and P21 were deeply anaesthetized with pentobarbital (50 mg kg⁻¹) and decapitated; brains were removed and immersed in cold (4 °C) ‘cutting’ solution containing (in mM): 234 sucrose, 11 glucose, 24 NaHCO₃, 2.5 KCl, 1.25 NaH₂PO₄, 10 MgSO₄ and 0.5 CaCl₂, gassed with 95% O₂/5% CO₂. Coronal slices (300 μ m) were cut from somatosensory cortex (parietal area 1) with a vibratome and then incubated in oxygenated artificial cerebrospinal fluid containing (in mM): 126 NaCl, 26 NaHCO₃, 2.5 KCl, 1.25 NaH₂PO₄, 2 MgSO₄, 2 CaCl₂ and 10 glucose, pH 7.4, initially at 32 °C for 1 h, and subsequently at 23–25 °C, before being transferred to the recording chamber and maintained at 32 °C. Recordings were obtained from layer V interneurons identified visually, which were easily distinguished from pyramidal neurons by their lacking a large emerging apical dendrite. Interneuron subtypes were identified by single AP waveforms, firing frequency, degree of spike frequency accommodation and dynamic range³, and by their GABAergic synaptic currents⁸. Depolarizing current pulses generated high-frequency, non-accommodating firing in FS cells. They lacked rebound spiking after hyperpolarizing current steps (Fig. 1c, inset) and had APs with prominent after-hyperpolarizations⁸. LTS cells generated a spike or a burst of spikes after hyperpolarizing current steps and had much lower firing frequencies and more spike frequency accommodation than FS cells⁸ (Fig. 1a, inset). Intracellular labelling with biocytin was used to confirm the interneuronal morphology in some cells as reported previously^{3,8} (data not shown, but see Supplementary Fig. S1). Experiments were performed in the whole-cell configuration of the patch-clamp technique. Electrodes (tip resistance 2–3 M Ω) were filled with an intracellular solution containing (in mM): potassium gluconate 70, KCl 70, NaCl 2, HEPES 10, EGTA 10, MgCl₂ 2; pH adjusted to 7.3 with KOH; 290 mOsm. In some experiments 10 mM BAPTA was substituted for 10 mM EGTA. Drugs were delivered with a local perfusion system composed of multiple fine tubes ending in a common outlet tube, positioned in proximity (about 250 μ m) to the recorded neuron. Experiments were performed in the presence of the ionotropic glutamate receptor blockers 6,7-dinitroquinoxaline-2,3-dione (10 μ M) and D,L-2-amino-5-phosphonvaleric acid (100 μ M) in the bath and local perfusate. In some experiments the GABA_A receptor blocker gabazine (10 μ M) was included in the perfusate. Signals were amplified, with a Multiclamp 700A patch-clamp amplifier (Axon Instruments), sampled at 20 kHz and filtered at 10 kHz unless otherwise noted. A Digidata 1320 digitizer and PClamp9 (Axon Instruments) as well as locally generated software (J.R.H.) were used for data acquisition and analysis and for the generation of stimuli. The membrane conductance tests used involved measurements of responses to small current injections (–15 to 30 pA, 250 ms, 0.2 Hz). SSI-inducing stimuli consisted of ten trains of either 10-Hz or 50-Hz APs (60 APs per train), evoked every 20 s. Results are presented as means $\pm$ s.e.m. Unless otherwise noted, the paired Student's *t*-test was used to compare control data with those obtained in the same neurons after drug applications or 5–8 min after SSI-inducing stimuli. Differences were considered significant at $p < 0.05$.

Immunohistochemistry and histology

Biocytin (0.1–0.05%; Sigma) was included in the internal solution to fill neurons during electrophysiological recordings. Slices were subsequently fixed overnight in 4% paraformaldehyde in phosphate buffer (pH 7.4) at 4 °C, cryoprotected in 30% sucrose in 0.1 M phosphate buffer and sectioned (50 μ m thickness) in a sliding microtome (MICROM International GmbH, Walldorf, Germany). Sections were treated twice in 50% ethanol for 10 min each time and blocked with 10% normal goat serum in PBS for 1 h, rinsed with PBS and incubated overnight with anti-CCK antibody (5 mg ml⁻¹, diluted 1:1,000 in PBS containing 0.5% Triton X-100). For CCK immunostaining we used a monoclonal antibody raised against CCK/gastrin (no. 28.2 monoclonal antibody) that was

kindly provided by G. Ohning (CURE/Digestive Diseases Research Center, Antibody/RIA Core, University of California, Los Angeles). Sections were rinsed twice in PBS and incubated for 2 h in the secondary antibody (goat anti-mouse coupled to fluorescein isothiocyanate, dilution 1:100; Molecular Probes), diluted in PBS plus 0.5% Triton X-100. Sections were rinsed in PBS and incubated for 1 h in Texas red Avidin D (diluted 1:1000; Vector Laboratories). Fluorescent biocytin-filled neurons and CCK immunoreactivity were then observed with a laser confocal microscope (Zeiss LSM510) and images were acquired.

Received 30 March; accepted 11 August 2004; doi:10.1038/nature02913.

- McBain, C. J. & Fisahn, A. Interneurons unbound. *Nature Rev. Neurosci.* **2**, 11–23 (2001).
- Freund, T. F. Interneuron diversity series: Rhythm and mood in perisomatic inhibition. *Trends Neurosci.* **26**, 489–495 (2003).
- Bacci, A., Huguenard, J. R. & Prince, D. A. Functional autaptic neurotransmission in fast-spiking interneurons: a novel form of feedback inhibition in the neocortex. *J. Neurosci.* **23**, 859–866 (2003).
- Thomson, A. M. & Deuchars, J. Synaptic interactions in neocortical local circuits: dual intracellular recordings *in vitro*. *Cereb. Cortex* **7**, 510–522 (1997).
- Tamas, G., Buhl, E. H. & Somogyi, P. Fast IPSPs elicited via multiple synaptic release sites by different types of GABAergic neurons in the cat visual cortex. *J. Physiol. (Lond.)* **500**, 715–738 (1997).
- Xiang, Z., Huguenard, J. R. & Prince, D. A. Synaptic inhibition of pyramidal cells evoked by different interneuronal subtypes in layer V of rat visual cortex. *J. Neurophysiol.* **88**, 740–750 (2002).
- Williams, S. R. & Stuart, G. J. Dependence of EPSP efficacy on synapse location in neocortical pyramidal neurons. *Science* **295**, 1907–1910 (2002).
- Bacci, A., Rudolph, U., Huguenard, J. R. & Prince, D. A. Major differences in inhibitory synaptic transmission onto two neocortical interneuron subclasses. *J. Neurosci.* **23**, 9664–9674 (2003).
- Beierlein, M., Gibson, J. R. & Connors, B. W. A network of electrically coupled interneurons drives synchronized inhibition in neocortex. *Nature Neurosci.* **3**, 904–910 (2000).
- Yamada, M., Inanobe, A. & Kurachi, Y. G protein regulation of potassium ion channels. *Pharmacol. Rev.* **50**, 723–760 (1998).
- Pouzat, C. & Marty, A. Somatic recording of GABAergic autoreceptor current in cerebellar stellate and basket cells. *J. Neurosci.* **19**, 1675–1690 (1999).
- Freund, T. F., Katona, I. & Piomelli, D. Role of endogenous cannabinoids in synaptic signaling. *Physiol. Rev.* **83**, 1017–1066 (2003).
- Di Marzo, V. *et al.* Formation and inactivation of endogenous cannabinoid anandamide in central neurons. *Nature* **372**, 686–691 (1994).
- Stella, N., Schweitzer, P. & Piomelli, D. A second endogenous cannabinoid that modulates long-term potentiation. *Nature* **388**, 773–778 (1997).
- Mackie, K., Lai, Y., Westenbroek, R. & Mitchell, R. Cannabinoids activate an inwardly rectifying potassium conductance and inhibit Q-type calcium currents in ArT20 cells transfected with rat brain cannabinoid receptor. *J. Neurosci.* **15**, 6552–6561 (1995).
- Kreitzer, A. C. & Regehr, W. G. Retrograde inhibition of presynaptic calcium influx by endogenous cannabinoids at excitatory synapses onto Purkinje cells. *Neuron* **29**, 717–727 (2001).
- Wilson, R. I. & Nicoll, R. A. Endogenous cannabinoids mediate retrograde signalling at hippocampal synapses. *Nature* **410**, 588–592 (2001).
- Ohno-Shosaku, T., Maejima, T. & Kano, M. Endogenous cannabinoids mediate retrograde signals from depolarized postsynaptic neurons to presynaptic terminals. *Neuron* **29**, 729–738 (2001).
- Losonczy, A., Biro, A. A. & Nusser, Z. Persistently active cannabinoid receptors mute a subpopulation of hippocampal interneurons. *Proc. Natl Acad. Sci. USA* **101**, 1362–1367 (2004).
- Trettel, J., Fortin, D. A. & Levine, E. S. Endocannabinoid signaling selectively targets perisomatic inhibitory inputs to pyramidal neurons in juvenile mouse neocortex. *J. Physiol. (Lond.)* **556**, 95–107 (2004).
- Piomelli, D. The molecular logic of endocannabinoid signalling. *Nature Rev. Neurosci.* **4**, 873–884 (2003).
- Diana, M. A. & Marty, A. Characterization of depolarization-induced suppression of inhibition using paired interneuron–Purkinje cell recordings. *J. Neurosci.* **23**, 5906–5918 (2003).
- Kreitzer, A. C., Carter, A. G. & Regehr, W. G. Inhibition of interneuron firing extends the spread of endocannabinoid signaling in the cerebellum. *Neuron* **34**, 787–796 (2002).
- Chevalayre, V. & Castillo, P. E. Heterosynaptic LTD of hippocampal GABAergic synapses: a novel role of endocannabinoids in regulating excitability. *Neuron* **38**, 461–472 (2003).
- Cravatt, B. F. *et al.* Supersensitivity to anandamide and enhanced endogenous cannabinoid signaling in mice lacking fatty acid amide hydrolase. *Proc. Natl Acad. Sci. USA* **98**, 9371–9376 (2001).
- Wischmeyer, E., Doring, F. & Karschin, A. Acute suppression of inwardly rectifying Kir2.1 channels by direct tyrosine kinase phosphorylation. *J. Biol. Chem.* **273**, 34063–34068 (1998).
- Galarreta, M. & Hestrin, S. A network of fast-spiking cells in the neocortex connected by electrical synapses. *Nature* **402**, 72–75 (1999).
- Gibson, J. R., Beierlein, M. & Connors, B. W. Two networks of electrically coupled inhibitory neurons in neocortex. *Nature* **402**, 75–79 (1999).
- Tamas, G., Somogyi, P. & Buhl, E. H. Differentially interconnected networks of GABAergic interneurons in the visual cortex of the cat. *J. Neurosci.* **18**, 4255–4270 (1998).
- Grenier, F., Timofeev, I. & Steriade, M. Neocortical very fast oscillations (ripples, 80–200 Hz) during seizures: intracellular correlates. *J. Neurophysiol.* **89**, 841–852 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank I. Parada for assistance during these experiments, and J. Krey for discussions. This work was supported by grants from the National Institute of Neurological Diseases and Stroke and the Pimley Research Fund.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.A.P. (daprin@stanford.edu).

Conserved mechanisms of glucose sensing and regulation by *Drosophila corpora cardiaca* cells

Seung K. Kim¹ & Eric J. Rulifson²

¹Departments of Developmental Biology and Medicine (Oncology Division), Stanford University School of Medicine, Beckman Center B300, Stanford, California 94305-5329, USA

²Department of Cell and Developmental Biology, University of Pennsylvania School of Medicine, 1214 BRB II/III, 421 Curie Boulevard, Philadelphia, Pennsylvania 19104-6058, USA

Antagonistic activities of glucagon and insulin control metabolism in mammals, and disruption of this balance underlies diabetes pathogenesis. Insulin-producing cells (IPCs) in the brain of insects such as *Drosophila* also regulate serum glucose^{1,2}, but it remains unclear whether insulin is the sole hormonal regulator of glucose homeostasis and whether mechanisms of glucose-sensing and response in IPCs resemble those in pancreatic islets. Here we show, by targeted cell ablation, that *Drosophila corpora cardiaca* (CC) cells^{3–5} of the ring gland are also essential for larval glucose homeostasis. Unlike IPCs, CC cells express *Drosophila* cognates of sulphonylurea receptor (Sur) and potassium channel (Ir), proteins that comprise ATP-sensitive potassium channels regulating hormone secretion by islets and other mammalian glucose-sensing cells^{6–8}. They also produce adipokinetic hormone, a polypeptide with glucagon-like functions. Glucose regulation by CC cells is impaired by exposure to sulphonylureas, drugs that target the Sur subunit. Furthermore, ubiquitous expression of an *akh* transgene reverses the effect of CC ablation on serum glucose. Thus, *Drosophila* CC cells are crucial regulators of glucose homeostasis and they use glucose-sensing and response mechanisms similar to islet cells.

Insect corpora cardiaca (CC) are clusters of endocrine cells in the ring gland adjacent to the prothoracic gland and corpus allatum. A principal CC product is adipokinetic hormone (AKH), a polypeptide that mobilizes stored macromolecular energy reserves to sustain energy-consuming activities, such as crawling and flight. AKH is similar to mammalian glucagon⁴; like glucagon in pancreatic islet α -cells, AKH is synthesized as a pre-prohormone, processed, and stored in dense core vesicles^{4,5}. Like mammalian glucagon activity in liver, AKH has been shown to bind a G-protein-coupled transmembrane receptor⁹ and to increase lipolysis, glycogenolysis and production of trehalose in the insect fat body, a storage organ for lipid and glycogen^{3,4}.

Previous studies of AKH microinjection and ring gland transplantation in locusts and other insects^{3,4} suggest that AKH is sufficient to increase haemolymph glucose concentrations, but have not yet shown a requirement for AKH in glucose homeostasis. To examine phenotypes resulting from CC cell ablation and AKH deficiency, we used a 1,000-base-pair DNA segment derived from sequences immediately 5' of the *Drosophila akh* gene (see Supplementary Information) to drive the expression of the transcriptional trans-activator GAL4 (Fig. 1a) in CC cells. The *akh*-GAL4 construct, when crossed with a *UAS-mCD8GFP* (membrane-tethered green fluorescent protein, mGFP) reporter line, directed a GFP expression pattern that reflected endogenous *akh* expression in the ring gland corpora cardiaca of third-instar larvae (Fig. 1a, b). Using *in situ* hybridizations, we established that embryonic *akh* messenger RNA expression initiates in cells of the presumptive CC anlage (Fig. 2a) and that in later larval stages it is maintained only in CC cells (Figs 1a, b and 2b, c). To assess the role of the CC as an endocrine regulator of haemolymph glucose concentrations, we

used *akh-GAL4* lines to express the cell death factor Reaper (ref. 10) in *akh*-expressing CC cells. This resulted in the ablation of only CC cells at high efficiency: in more than 96% of newly hatched first-instar larvae ($n = 40$) harbouring *akh-GAL4*, *UAS-Reaper* and *UAS-mCD8GFP*, no mGFP-labelled CC cells were detected. In contrast, mGFP was detected in CC cells within all control larvae at the same stage (data not shown), and at later stages (Figs 1b and 2b). In *Drosophila*, haemolymph glucose is composed of trehalose (a disaccharide of glucose) and monomeric free glucose, and the combined circulating concentration of these (hereafter referred to as total haemolymph glucose) is maintained in a narrow range for a given feeding condition (see Methods). Ablation of *akh*-expressing CC cells in larvae raised on dextrose-supplemented medium decreased the mean total haemolymph glucose (Fig. 1g) and trehalose by 50% ($1,509 \pm 188.5$ versus 688 ± 165 mg dl⁻¹; $P < 0.001$, $n = 7$), an effect similar to that recently reported by others¹¹. CC cell deficiency did not result in discernible growth reduction, developmental delay or lethality, phenotypes that arise

after the ablation of IPCs in the brain. Thus, like glucagon, AKH is an essential regulator of energy metabolism but might be dispensable for developmental growth control¹².

To test whether AKH activity alone could account for the glucose-regulating action of CC cells, we measured the ability of an *akh* transgene with ubiquitous expression from a heat shock promoter to reverse the effect of CC ablation (see Supplementary Information). Lower haemolymph glucose concentrations resulting from CC ablation were partly restored by the ubiquitous expression of an *akh* transgene (Fig. 1g). Thus, bioactive AKH from the *akh* transgene might be produced in target tissues, as has been shown for transgene-encoded neuropeptides such as *Drosophila* insulin (ref. 13). These data indicate that AKH is an essential regulator of haemolymph carbohydrate concentrations in *Drosophila*. We suggest that the hyperglycaemic effects of AKH counter-regulate the activity of other systemic hormones such as insulin and that these antagonistic activities might refine the levels of circulating energy to match systemic energy requirements. If so, we postulated

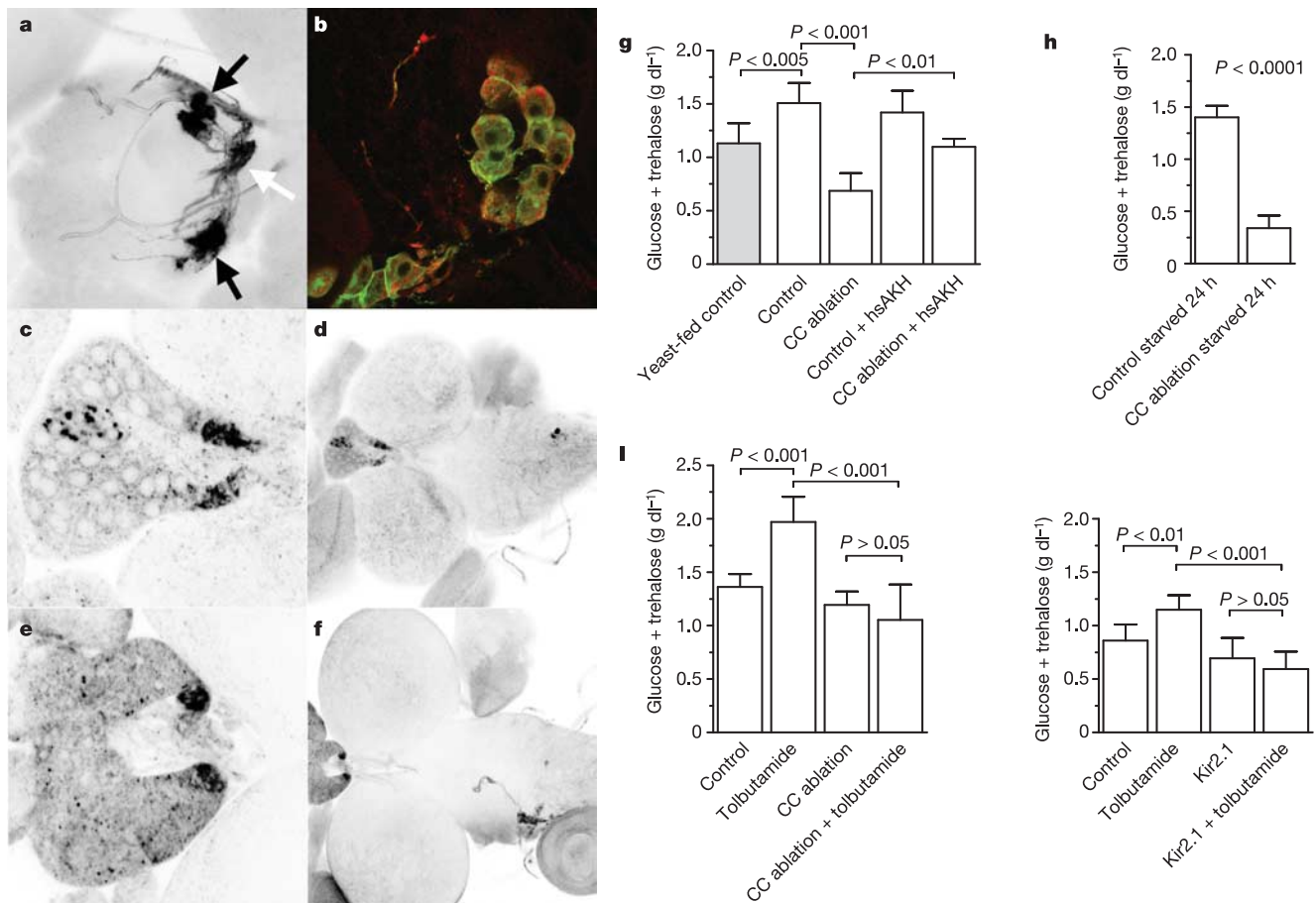


Figure 1 Haemolymph glucose regulation by K_{ATP} channel-expressing cells in the corpora cardiaca. Anterior is to the left in **a–f**. **a**, *akh-GAL4* directs mGFP expression (black) in third-instar corpora cardiaca cells (black arrows) and their projections to the heart (white arrow). **b**, Merged picture showing corpora cardiaca cells simultaneously labelled by *akh-GAL4*-directed expression of mCD8 GFP (green) and by antiserum that detects AKH protein (red). Strands to the left of the cell bodies are projections to the prothoracic gland. **c**, *In situ* hybridization revealing *Sur* expression in the paired corpora cardiaca and in the corpus allatum. **d**, *In situ* hybridization revealing lack of detectable *Sur* expression in the protocerebral lobes adjacent to the ring gland containing *Drosophila* IPCs. **e**, *In situ* hybridization revealing *Ir* expression in the paired corpora cardiaca. **f**, *In situ* hybridization revealing lack of *Ir* expression in the protocerebral lobes adjacent to the ring gland

containing IPCs. **g–i**, Haemolymph carbohydrate in *Drosophila* larvae is regulated by AKH and the sulphonylurea receptor complex in AKH-producing corpora cardiaca cells. Pairwise ANOVA tests and P values are indicated by brackets in **g–i**. Data from yeast-fed larvae. Numbers of samples per condition were $n = 7$ (**g**) and $n = 6$ (**h**, **i**). Results are means $\pm$ s.d. in all graphs. **g**, AKH is required for regulation of haemolymph glucose + trehalose concentrations in third-instar larvae on standard dextrose-supplemented medium. Mean concentrations of haemolymph glucose + trehalose in control larvae fed with standard dextrose medium (control) were higher than those in yeast-fed controls (yeast-fed). **h**, Starvation exacerbates the hypoglycaemic effect of CC cell ablation. **i**, Effect of ablating *Akh*-producing corpora cardiaca (CC) cells or of Kir2.1 expression in CC cells on larval tolbutamide responses.

that the negative energy balance accompanying starvation might worsen the hypoglycaemic effects of AKH deficiency. In comparison with starved control larvae, total haemolymph glucose was decreased by 75% in starving larvae after CC cell ablation (Fig. 1h). Thus, starvation increased the severity of hypoglycaemia in animals lacking CC cells, indicating that AKH might be required for the compensatory mechanisms that maintain circulating glucose during periods of food deprivation in *Drosophila* larvae.

Labelling of CC cell processes with mGFP and an antibody against AKH revealed that AKH-producing cells extend processes that terminate on the heart and on the prothoracic gland compartment of the ring gland (Fig. 2b, c; Supplementary Information). On the surface of the heart, CC cell processes have extensive contact with axons that project from insulin-producing cells from the brain (Fig. 2b). Labelling of CC cell processes with mGFP and an antibody against AKH revealed localization of AKH within the processes that contact the IPCs, and AKH peptide on the processes contacting the heart (Fig. 2d). These results indicate that the heart surface is the principal site of AKH release into the circulating haemolymph. Thus, like glucagon-producing cells in mammalian islets and brain, AKH-producing CC cells in the *Drosophila* ring gland have direct systemic vascular access, which is consistent with their role as endocrine regulators of metabolism.

ATP-sensitive potassium (K_{ATP}) channels regulate neuroendocrine cell function in organs such as the mammalian pancreas and brain^{6,7}, and we examined whether K_{ATP} functions regulate CC cell activity. K_{ATP} channels are heteromeric protein complexes composed of sulphonylurea receptor (Sur) and inward-rectifying potassium channel (Ir; also called Kir) subunits. An ATP-binding domain

in the Ir subunit regulates K_{ATP} channel activity, allowing these channels to serve as cellular energy sensors, opening or closing in response to the intracellular ADP/ATP ratio, thus influencing membrane potential and subsequent calcium currents that regulate hormone secretion. Using mRNA *in situ* hybridization (Fig. 1c–f), we showed that larval CC cells expressed *Sur* (ref. 14) and *Ir* (ref. 15), which have sequence similarity to mammalian Sur1 and Kir6 proteins, respectively. Expression of *Sur* or *Ir* was not detected in the larval brain IPCs, another group of cells known to regulate haemolymph glucose (Fig. 1d, f; refs 1, 2). *Drosophila* Sur has been shown to be sufficient to allow K^+ currents that polarize membrane potentials¹⁴. *Drosophila* Ir was demonstrated to evoke an inwardly rectifying K^+ current¹⁵.

We then tested whether increased haemolymph glucose concentrations might result from excess AKH secretion brought about by sulphonylurea inhibition of the Sur and K^+ -dependent depolarization of CC cells. Glyburide and tolbutamide are representative members of the two major classes of sulphonylureas⁸. These drugs promote the closure of K_{ATP} channels and cellular depolarization, thereby regulating secretion in mammalian neuroendocrine cells^{16,17}. For example, sulphonylureas stimulate glucagon secretion in diabetic patients¹⁸. Glyburide has previously been shown to inhibit *Drosophila* Sur-mediated outward K^+ currents, resulting in the depolarization of cell potential¹⁴. Exposure of feeding third-instar larvae to glyburide mixed in yeast paste (standard dextrose medium did not permit drug delivery; see Methods) produced a 10% increase in mean total haemolymph glucose concentration compared with controls ($2,110 \pm 300.3$ versus $1,915 \pm 209.9$ mg dl⁻¹; $P = 0.025$, $n = 7$). Exposure of larvae to

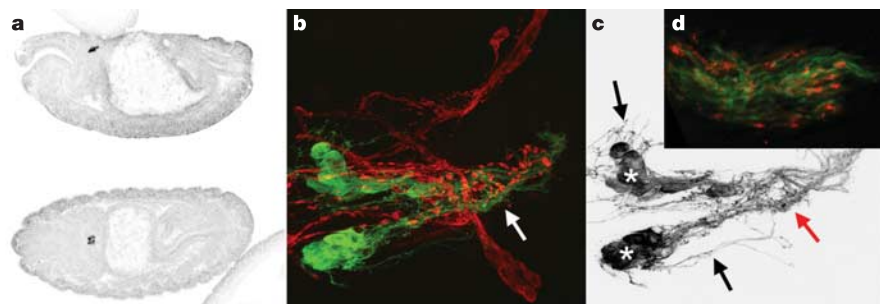


Figure 2 *akh* gene expression and morphology of *akh*-expressing cells. Anterior is to the left in all panels. **a**, Sagittal (top) and dorsal (bottom) views of stage-16 embryos after *in situ* hybridization with antisense probes specific for *akh* mRNA. **b**, Merged confocal microscopy image of *akh*-expressing cells expressing mCD8 (green) in contact with IPCs stained by anti-insulin antisera (red). There is extensive contact (yellow puncta) between

IPC axons and CC cell projections overlying the heart (arrow). **c**, Detail of CC cell bodies (asterisks) and their projections to the heart (red arrow) and projections within the ring gland prothoracic gland (black arrows). **d**, Labelling of mCD8 GFP-expressing CC projections (green) to the heart with anti-AKH antisera (red).

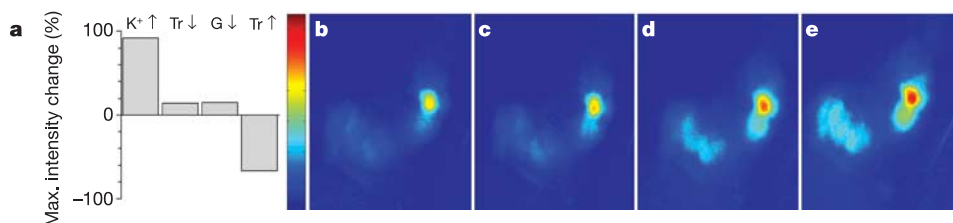


Figure 3 Changes in fluorescence intensity in *akh*-expressing cells producing camgaroo-2 (cg-2). **a**, Percentage change in emission intensity after potassium depolarization ($K^+ \uparrow$), trehalose decrease (Tr $\downarrow$), glucose decrease (G $\downarrow$) and trehalose increase (Tr $\uparrow$). **b–e**, Time course of increasing cg-2 emission after decrease in trehalose concentration

in vitro. Images were captured at 0 min (**b**), 1 min (**c**), 4 min (**d**) and 7 min (**e**), respectively, after decrease in trehalose concentration. A linear colour scale (red shows highest intensity, blue lowest) is shown at the left of **b**.

tolbutamide had a greater effect, producing a 40% increase in mean total haemolymph glucose (Fig. 1i) and we used tolbutamide in subsequent studies. Average haemolymph glucose concentrations were generally decreased in animals fed with yeast paste compared with animals fed with standard dextrose medium (Fig. 1g), and this might have accentuated the hyperglycaemic effect of sulphonylureas administered in yeast paste. Moreover, the hypoglycaemic effect induced by CC cell ablation (or hyperpolarization; see below) seemed attenuated in yeast-fed animals (Fig. 1g, i), further supporting our hypothesis that requirements for AKH might be altered by manipulating feeding conditions.

To test the hypothesis that Sur and Ir function in the CC to regulate haemolymph glucose concentrations in *Drosophila*, we ablated CC cells in larvae fed with tolbutamide. Ablation of the CC cells using *Akh-GAL4* and *UAS-Reaper* blocked the hyperglycaemic effect of tolbutamide (Fig. 1i), indicating that CC cells must be present to support the hyperglycaemic action of tolbutamide. To determine whether the hyperglycaemic effect of tolbutamide resulted from Sur and Ir-mediated depolarization of CC cells, we hyperpolarized membrane potential in CC cells, in the presence and absence of tolbutamide. Kir2.1 is a human K⁺ channel that evokes an outward K⁺ current, independently of ATP regulation, and has previously been used¹⁹ to impair cellular depolarization *in vivo* in *Drosophila* by inducing persistent outward K⁺ current and a hyperpolarized resting potential. One indication that AKH release by CC cells requires membrane depolarization and might be regulated by K⁺-channel-dependent membrane potential comes from our observation that, on standard dextrose medium, third-instar larvae expressing Kir2.1 in CC cells had a 23% decrease in mean haemolymph glucose concentration, compared with controls (693.7 ± 192.9 versus 1,106 ± 105.5 mg dl⁻¹; *P* < 0.01, *n* = 6). Expression of the Kir2.1 channel in CC cells prevented the hyperglycaemic effect of tolbutamide (Fig. 1i), indicating that K⁺-channel-dependent CC cell depolarization resulted from exposure to sulphonylurea. Together, these pharmacological and genetic data support the view that K_{ATP} channel activity in CC cells governs AKH release, thereby controlling concentrations of circulating glucose in *Drosophila*.

In pancreatic α-cells, hypoglycaemia stimulates increased intracellular calcium concentrations promoting glucagon secretion, whereas hyperglycaemia inhibits these responses^{20–22}. To test whether *Drosophila* CC cells sense glucose changes and, like pancreatic α-cells, modulate intracellular calcium concentrations, we marked CC cells with fluorescent transgene-encoded calcium sensors ('camgaros'). The fluorescence intensity of camgaros increases in response to elevated intracellular calcium concentration, an effect used previously to measure cytoplasmic calcium transients in depolarized *Drosophila* neurons²³. Elevation of cytoplasmic calcium concentration after CC cell depolarization stimulates AKH secretion^{24,25}; thus, in our experiments we used elevated intracellular calcium concentration in CC cells as an indicator of AKH secretion. Fluorescence of camgaroo-2 (cg-2) in cultured CC cells (Fig. 3a) increased as extracellular trehalose or glucose concentration decreased (Fig. 3b–e). Direct CC cell depolarization with increased extracellular potassium concentration similarly led to increased cg-2 fluorescence (Fig. 3a). In contrast, fluorescence in cg-2-labelled CC cells decreased as extracellular trehalose concentration increased (Fig. 3a). These results corroborate previous studies of locust CC cells showing that decreases in extracellular trehalose or glucose concentration stimulated AKH secretion²⁶. *Drosophila* CC cells express the enzyme trehalase (ref. 27), raising the possibility that the sensing of trehalose by CC cells involves the hydrolysis of trehalose to glucose, a view also supported by similar effects of trehalose and glucose in our studies *in vitro*. Thus, hypoglycaemic sensing in CC cells leads to increased concentrations of the intracellular second messenger calcium, a signal for subsequent regulated exocytosis of AKH^{25,28}—a mechanism similar

to those regulating glucagon secretion by mammalian pancreatic α-cells.

Thus, there are remarkable parallels in endocrine cell functions that ensure the supply of circulating glucose in Diptera and in mammals. On the basis of these parallels, we speculate that insect CC cells and mammalian neuroendocrine cells that regulate metabolism might have arisen from an ancestral energy-sensing cell. If so, we speculate further that pancreatic islet cells, including β-cells, might have evolved from an ancient α-cell. Similarly to pancreatic islets, insect CC cells might delaminate from embryonic epithelial cells that give rise to both gut and neuroendocrine structures (ref. 29, 30). Thus, common mechanisms might regulate the development of CC and pancreatic islet cells. Understanding CC cell development could therefore accelerate the discovery of cell-replacement therapies for type 1 diabetes mellitus. This *Drosophila* model might also serve to elucidate the mechanisms that control stimulus–secretion coupling in CC cells, and hence the biology of hypoglycaemia. Moreover, the sensitivity of CC cells to drugs commonly prescribed for disorders such as type 2 diabetes indicates that *Drosophila* might provide a model system for the discovery of pharmacological agents to treat human endocrine diseases. □

Methods

Antibody generation

To generate antisera specific for the AKH protein, PCR was used to amplify an internal fragment of the *akh* cDNA encoding the 79-amino-acid peptide sequence MNPKS...CKHRE as described in the Supplementary Information. This DNA fragment was then cloned into pCRII, sequenced, and then excised by *Xba*I and *Eco*R1 digestion and ligated into the glutathione S-transferase (GST) fusion protein plasmid pGEX4T-2 (Pharmacia), which had been linearized by digestion with *Xba*I and *Eco*R1 restriction enzymes. GST fusion protein was produced and purified as described by the vendor and used as immunogen in rabbits. The antiserum was affinity-purified against the immunogen, which was covalently immobilized to Sepharose. The antiserum reacted with bands on an immunoblot with apparent molecular mass corresponding to the predicted size of the AKH–GST fusion protein.

Trehalose and glucose measurements

Haemolymph samples from third-instar larvae were collected, and concentrations of trehalose and glucose were measured as described previously¹, with the following modifications. Haemolymph samples were collected from groups of two or three larvae with a micropipette (Transferpette; Brand, Germany) with a fixed volume setting. In 96-well plates, 0.2-μl haemolymph samples and standards were added to 120 μl of Infinity Glucose Reagent, adjusted to pH 6.8 (ThermoDMA) with pig kidney trehalase at 10 μl per 5 ml (Sigma). Standard curves were generated for trehalose and glucose concentration for every experimental trial; the assay was linear over a 0–2,500 mg dl⁻¹ range. A linear regression of concentration samples over this range was used to find actual average concentration values for haemolymph samples. One-way ANOVA with a pairwise Bonferroni test was applied to each experimental trial to determine significance. Results are presented as means ± s.d. After CC ablation in dextrose-medium-fed larvae, we did not detect changes in concentrations of free glucose in larval haemolymph. All results were repeated at least twice. The contribution of background NADH in haemolymph to glucose concentrations in the assay was found to be insignificant when identical values were obtained in parallel with an alternative assay that did not use NADH as an endpoint (Liquid Glucose Reagent Set; Pointe Scientific) (E. Bustamante, G. McLean & S. Kim, unpublished observations).

Larval methods

The sizes of staged larvae at first instar and wandering third instar were quantified as described previously. Egg laying and larval growth were performed routinely in dextrose medium. Dextrose medium is made by adding 128.4 g anhydrous dextrose (VWR Scientific), 9.3 g agar (Fine Ground USP; Morehead), 61.2 g corn meal (ICN) and 32.4 g yeast (Lakes States) to 1.1 l distilled water, prewarmed to 70 °C. After heating of this mixture to 97 °C, the medium was cooled to 70 °C, then supplemented with a 10% (w/v) solution of methyl-*p*-hydroxybenzoate (an antifungal reagent; ICN) and cooled in small plastic vials. Third-instar larvae feeding in dextrose medium vials were washed briefly in PBS, then placed in vials with water-soaked filter paper for starvation experiments. For sulphonylurea experiments, washed larvae were placed on grape-juice agar plates supplemented with 1 g of yeast paste (with the consistency of smooth peanut butter) made by thorough mixing of about 5 ml water and 2.5 g yeast (Fleischmann's yeast; Burns Philp Food Ltd, Lasalle, Canada). With 1 g yeast paste we mixed 120 mg tolbutamide (Sigma) or 80 mg glyburide (Sigma) in some experiments. Neither of these drugs was consistently suspended or deliverable in standard dextrose medium. After 12 h at 21–22 °C, feeding larvae were harvested and analysed as described in the text.

Received 12 May; accepted 27 July 2004; doi:10.1038/nature02897.

- Rulifson, E. J., Kim, S. K. & Nusse, R. Ablation of insulin-producing neurons in flies: growth and diabetic phenotypes. *Science* **296**, 1118–1120 (2002).
- Ikeya, T., Galic, M., Belawat, P., Nairz, K. & Hafen, E. Nutrient-dependent expression of insulin-like peptides from neuroendocrine cells in the CNS contributes to growth regulation in *Drosophila*. *Curr. Biol.* **12**, 1293–1300 (2002).
- Veelaert, D., Schoofs, L. & De Loof, A. Peptidergic control of the corpus cardiacum-corpora allata complex of locusts. *Int. Rev. Cytol.* **182**, 249–302 (1998).
- Van der Horst, D. J. Insect adipokinetic hormones: release and integration of flight energy metabolism. *Comp. Biochem. Phys. B* **136**, 217–226 (2003).
- Noyes, B. E., Katz, F. N. & Schaffer, M. H. Identification and expression of the *Drosophila* adipokinetic hormone gene. *Mol. Cell. Endocrinol.* **109**, 133–141 (1995).
- Aguilar-Bryan, L. et al. Cloning of the beta cell high-affinity sulfonylurea receptor: a regulator of insulin secretion. *Science* **268**, 423–426 (1995).
- Miki, T. et al. ATP-sensitive K⁺ channels in the hypothalamus are essential for the maintenance of glucose homeostasis. *Nature Neurosci.* **4**, 507–512 (2001).
- Seino, S. & Miki, T. Physiological and pathophysiological roles of ATP-sensitive K⁺ channels. *Prog. Biophys. Mol. Biol.* **81**, 133–176 (2003).
- Staubli, F. et al. Molecular identification of the insect adipokinetic hormone receptors. *Proc. Natl Acad. Sci. USA* **99**, 3446–3451 (2002).
- McNabb, S. L. et al. Disruption of a behavioral sequence by targeted death of peptidergic neurons in *Drosophila*. *Neuron* **19**, 813–823 (1997).
- Lee, G. & Park, J. H. Hemolymph sugar homeostasis and starvation-induced hyperactivity affected by genetic manipulations of the adipokinetic hormone-encoding gene in *Drosophila melanogaster*. *Genetics* **167**, 311–323 (2004).
- Gelling, R. W. et al. Lower blood glucose, hyperglucagonemia, and pancreatic alpha cell hyperplasia in glucagon receptor knockout mice. *Proc. Natl Acad. Sci. USA* **100**, 1438–1443 (2003).
- Brogiolo, W. et al. An evolutionarily conserved function of the *Drosophila* insulin receptor and insulin-like peptides in growth control. *Curr. Biol.* **11**, 213–221 (2001).
- Nasonkin, I. et al. A novel sulfonylurea receptor family member expressed in the embryonic *Drosophila* dorsal vessel and tracheal system. *J. Biol. Chem.* **274**, 29420–29425 (1999).
- Döring, E., Wischmeyer, E., Kuhnlein, R. P., Jäckle, H. & Karschin, A. Inwardly rectifying K⁺ (Kir) channels in *Drosophila*. A crucial role of cellular milieu factors Kir channel function. *J. Biol. Chem.* **277**, 25554–25561 (2002).
- Sturgess, N. C., Ashford, M. L., Cook, D. L. & Hales, C. N. The sulfonylurea receptor may be an ATP-sensitive potassium channel. *Lancet* **ii**, 474–475 (1985).
- Trube, G., Rorsman, P. & Ohno-Shosaku, T. Opposite effects of tolbutamide and diazoxide on the ATP-dependent K⁺ channel in mouse pancreatic beta-cells. *Pflügers Arch.* **407**, 493–499 (1986).
- Ostergaard, T. et al. The insulin secretagogues glibenclamide and repaglinide do not influence growth hormone secretion in humans but stimulate glucagon secretion during profound insulin deficiency. *J. Clin. Endocrinol. Metab.* **89**, 297–302 (2004).
- Paradis, S., Sweeney, S. T. & Davis, G. W. Homeostatic control of presynaptic release is triggered by postsynaptic membrane depolarization. *Neuron* **30**, 737–747 (2001).
- Göpel, S. O. et al. Regulation of glucagon release in mouse-cells by KATP channels and inactivation of TTX-sensitive Na⁺ channels. *J. Physiol. (Lond.)* **528**, 509–520 (2000).
- Hoy, M. et al. Tolbutamide stimulates exocytosis of glucagon by inhibition of a mitochondrial-like ATP-sensitive K⁺ (KATP) conductance in rat pancreatic α -cells. *J. Physiol. (Lond.)* **527**, 109–120 (2000).
- Bokvist, K. et al. Characterisation of sulphonylurea and ATP-regulated K⁺ channels in rat pancreatic α -cells. *Pflügers Arch.* **438**, 428–436 (1999).
- Yu, D., Baird, G. S., Tsien, R. Y. & Davis, R. L. Detection of calcium transients in *Drosophila* mushroom body neurons with camguroo reporters. *J. Neurosci.* **23**, 64–72 (2003).
- Pannabecker, T. & Orchard, I. Regulation of adipokinetic hormone release from locust neuroendocrine tissue: participation of calcium and cyclic AMP. *Brain Res.* **423**, 13–22 (1987).
- Pannabecker, T. & Orchard, I. Ionic dependence of depolarization-mediated adipokinetic hormone release from the locust corpus cardiacum. *Brain Res.* **477**, 38–47 (1989).
- Passier, P., Vullings, H. G. B., Diederer, J. H. B. & Van der Horst, D. J. Trehalose inhibits the release of adipokinetic hormones from the corpus cardiacum in the African migratory locust, *Locusta migratoria*, at the level of the adipokinetic cells. *J. Endocrinol.* **153**, 299–305 (1997).
- Tomancak, P. et al. Berkeley *Drosophila* Genome Project (<http://www.fruitfly.org/cgi-bin/ex/insitu.pl>).
- Orchard, I. & Loughton, B. G. Is octopamine a transmitter mediating hormone release in insects? *J. Neurobiol.* **12**, 143–153 (1981).
- Copenhaver, P. F. & Taghert, P. H. Origins of the insect enteric nervous system: differentiation of the enteric ganglia from a neurogenic epithelium. *Development* **113**, 111–132 (1991).
- DeVelasco, B., Shen, J., Go, S. & Hartenstein, V. Embryonic development of the *Drosophila* corpus cardiacum, a neuroendocrine gland with similarity to the vertebrate pituitary, is controlled by *sine oculis* and *glass*. *Dev. Biol.* (in the press).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank B. Tsai for *Ir* and *Sur* cloning and *in situ* hybridizations, X. Gu for *akh* cloning and constructions, G. McLean for fly husbandry, E. Bustamante for glucose assays, M. Fish for P-element-mediated transformations, P. Calvert for quantitative analysis of calcium transients, and R. Nusse, G. Barsh, M. Krasnow, S. DiNardo and members of the Rulifson and Kim laboratories for suggestions on the manuscript. E.R. was supported by a grant from the Juvenile Diabetes Research Foundation (JDRF). S.K. was supported by grants from the JDRF, the Pew Charitable Trusts and the Verto Institute. E.R. and S.K. contributed equally to this study.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.K.K. (seungkim@cmgm.stanford.edu).

Spermatid differentiation requires the assembly of a cell polarity complex downstream of junctional adhesion molecule-C

Georgia Gliki¹, Klaus Ebnet², Michel Aurrand-Lions³, Beat A. Imhof³ & Ralf H. Adams¹

¹Vascular Development Laboratory, Cancer Research UK London Research Institute, London WC2A 3PX, UK

²Institute of Cell Biology, ZMBE, University Hospital Muenster, D-48149 Muenster, Germany

³University of Geneva, CMU, Department of Pathology, CH-1211 Geneva, Switzerland

During spermatogenesis in the mammalian testis, stem cells (spermatogonia) differentiate into spermatocytes, which subsequently undergo two consecutive meiotic divisions to give rise to haploid spermatids. These cells are initially round but progressively elongate, condense their nuclei, acquire flagellar and acrosomal structures, and shed a significant amount of their cytoplasm to form spermatozoa (the sperm cells) in a developmental cascade termed spermiogenesis^{1,2}. Defects in these processes will lead to a lack of mature sperm cells (azoospermia), which is a major cause of male infertility in the human population³. Here we report that a cell-surface protein of the immunoglobulin superfamily, junctional adhesion molecule-C (JAM-C), is critically required for the differentiation of round spermatids into spermatozoa in mice. We found that *Jam-C* is essential for the polarization of round spermatids, a function that we attribute to its role in the assembly of a cell polarity complex.

Members of the JAM family, which include JAM-A, JAM-B and JAM-C, are characterized by two extracellular Ig domains (mediating homophilic and heterophilic interactions), a single transmembrane region and a short cytoplasmic domain. Various studies have suggested that JAM proteins have important functions in the assembly of tight junctions in endothelial and epithelial cells, transendothelial migration of immune cells and in inflammatory processes^{4–8}. To study the function of *Jam-C* *in vivo*, we have disrupted the gene by homologous recombination in mice (for details see Methods and Supplementary Fig. 1). Although a major proportion of the *Jam-C*^{−/−} offspring died during postnatal development, about 40% of the mutants were viable. *Jam-C*-deficient males were infertile and failed to produce mature sperm cells (Supplementary Fig. 2). Dissected *Jam-C*^{−/−} testes were about 50% smaller than littermate controls and lacked differentiated elongated spermatids (Fig. 1a, c, d). Aided by a *LacZ* reporter cassette contained in our targeting construct, we observed prominent *Jam-C* expression at the sites of spermatogenesis in testis—the seminiferous tubules (Fig. 1a). Immunofluorescence on tissue sections showed JAM-C protein in most generations of spermatogenic cells, that is, in diploid pre-meiotic spermatocytes, haploid spermatids and the more differentiated elongated spermatids covering the luminal surface of the seminiferous tubules (Fig. 1b). JAM-C was widely distributed on round spermatogenic cells but a fraction of the protein was concentrated in the anterior area (Fig. 1e). As spermiogenesis proceeded, JAM-C became confined to the junctional plaques in the heads of elongated spermatids (Fig. 1f). Junctional plaques are specialized adhesion structures that anchor spermatids to the Sertoli cell epithelium, which provides spermatogenic cells with essential signals and nutrients. A second type of junctional structure in seminiferous tubules, the Sertoli–Sertoli tight junctions, showed no JAM-C signal (Fig. 1i; and Supplementary

Fig. 3). Instead, the JAM family members JAM-A and JAM-B were present in these junctions (Fig. 1j; and Supplementary Fig. 3).

JAM-B was also present in junctional plaques connecting Sertoli cells with round and elongated spermatids (Fig. 1g, h). Because JAM-B and JAM-C are counter-receptors with a function in the adhesion of immune cells^{9,10}, we explored the possibility that the two JAM proteins may mediate interactions between Sertoli cells and spermatids. Indeed, double immunofluorescence showed over-

lapping JAM-B and JAM-C signals near the lumen of seminiferous tubules (Fig. 1k). High magnification images suggest that JAM-B is presented by the Sertoli cell at the junctional plaque and may provide a partner for trans-interactions with JAM-C found on spermatids (Fig. 1l, m). Consequently, the contact between different JAM proteins at the junctional plaque may be essential for Sertoli cell-spermatid communication and spermatid differentiation.

As disruption of Sertoli cell tight junctions through the breakdown of the blood-testis barrier is a possible cause for male infertility^{1,11}, we examined this structure in the *Jam-C*^{-/-} testes. Consistent with the protein localization data, inactivation of the *Jam-C* gene did not lead to detectable changes in Sertoli-Sertoli tight junction proteins, tight junction ultrastructure or permeability of the blood-testes barrier as assessed by Evans blue perfusion (Supplementary Fig. 3; and data not shown). Having excluded a defective blood-testes barrier as the cause of male infertility, we next examined the *Jam-C*-deficient spermatogenic cells. Electron microscopic analysis of *Jam-C*^{-/-} round spermatids showed that they lacked acrosomal structures and all other morphological signs of polarization (Fig. 2a, b). All further steps of differentiation, such as the condensation of nuclei or the formation of flagella, failed in the absence of JAM-C (data not shown). Staining of spermatids for glycoproteins confirmed compromised acrosome formation in the absence of JAM-C (Fig. 2c-f).

Cytoskeletal elements at the Sertoli-spermatid interface have critical roles during normal spermiogenesis^{12,13}. Assembly of an F-actin scaffold at the Sertoli-spermatid junction requires heterotypic interactions between the cell adhesion molecules nectin-2

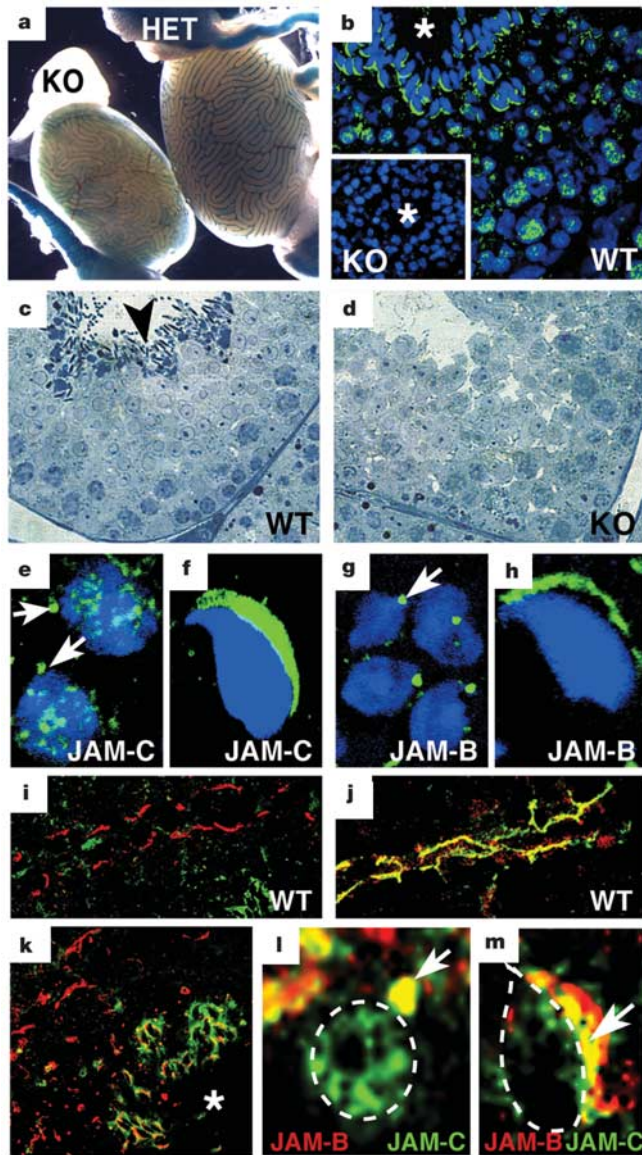


Figure 1 Defective spermiogenesis in *Jam-C*^{-/-} mutants. **a**, β -galactosidase staining of *Jam-C*^{+/-} (HET) and *Jam-C*^{-/-} (KO) testes. **b**, Cross-sections through wild-type (WT) and KO tubules (inset) stained with anti-JAM-C antibody (green). **c**, **d**, Toluidine-blue-stained sections of seminiferous tubules. Arrowhead indicates elongated spermatids in **c**. **e**–**m**, Confocal images of JAM-C and JAM-B. Nuclei are shown in blue (TO-PRO-3). JAM-C is present in junctional plaques of round (arrows in **e**) and elongated spermatids (**f**). JAM-B (green) signal at the Sertoli-spermatid interface of round (arrow in **g**) and elongated spermatids (**h**). Sertoli-Sertoli tight junctions containing ZO-1 (red) are devoid of JAM-C (green) (shown in **i**). Co-localization of JAM-B (green) and occludin (red) in Sertoli cell tight junctions (**j**). JAM-B (red) and JAM-C (green) signals overlap proximal to the lumen (asterisks) (shown in **k**). Co-localization of JAM-B (red) and JAM-C (green) at Sertoli-spermatid contact sites (arrows) (**l** and **m**). Nuclei are outlined.

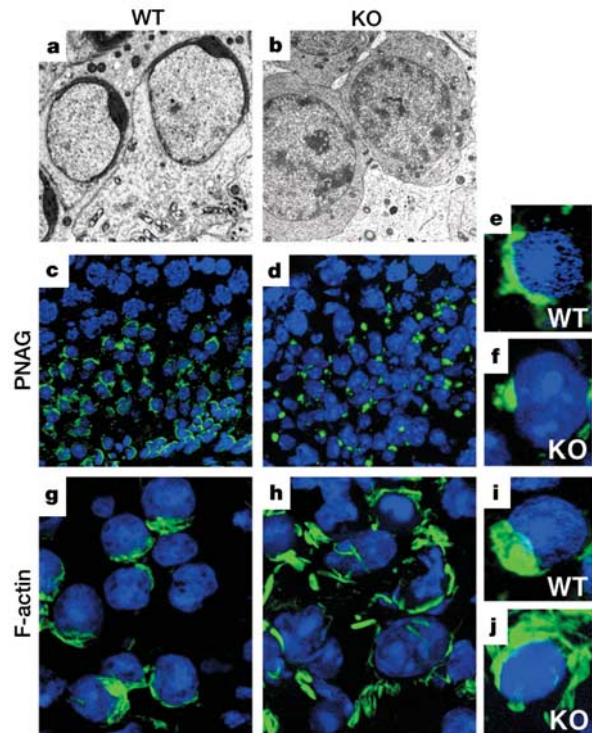


Figure 2 Differentiation arrest of *Jam-C*^{-/-} spermatids. **a**, **b**, Transmission electron micrographs of round spermatids. KO spermatids lack acrosomes (electron dense structures seen in **a**) and elongated nuclei. **c**–**j**, Confocal images of cross-sections from WT (**c**, **e**, **g**, **i**) and KO (**d**, **f**, **h**, **j**) seminiferous tubules. Nuclei are labelled with TO-PRO-3 (blue). Peanut agglutinin (PNAG) staining (**c**–**f**, green) labels the acrosomal cap in WT spermatids (**c**, **e**). Residual staining in *Jam-C*^{-/-} spermatids corresponds to Golgi structures (**d**, **f**). F-actin (**g**–**j**, phalloidin, green) is polarized at WT Sertoli-spermatid plaques (**g**, **i**) but is disorganized in mutants (**h**, **j**).

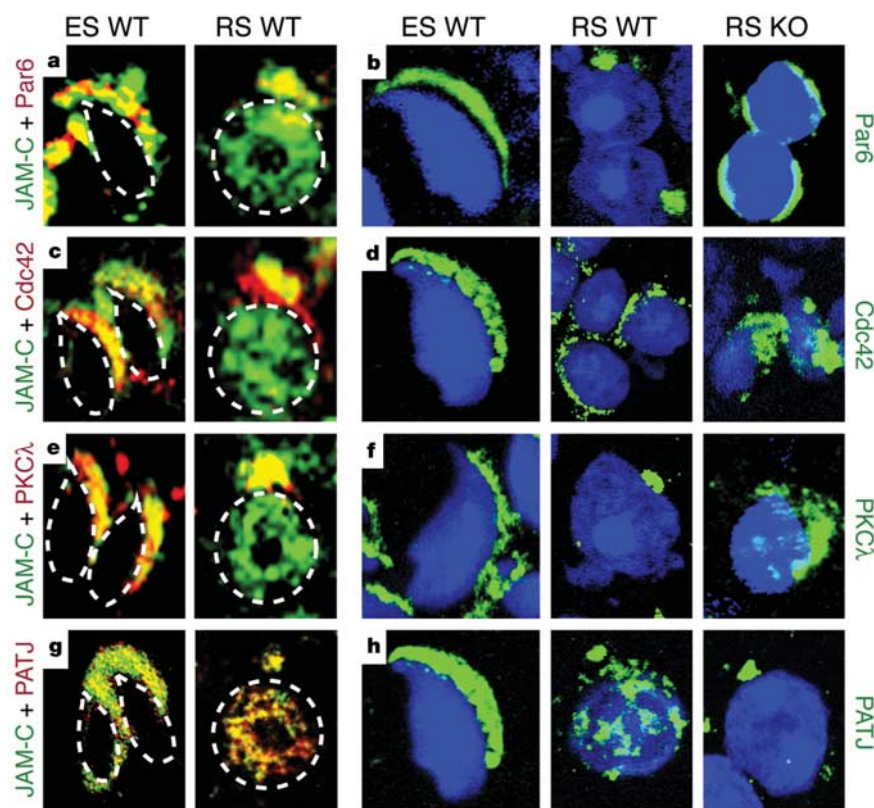


Figure 3 Compromised cell polarization in *Jam-C* KOs. **a–h**, Confocal images of round (RS) and elongated (ES) spermatids in sections through WT and KO seminiferous tubules. Nuclei are stained with TO-PRO-3 (shown in blue) or outlined by dotted lines. Primary antibodies were used as indicated (green and red). Note overlap of JAM-C with

Par6 (**a**), Cdc42 (**c**), PKC λ (**e**) or PATJ (**g**) in WT spermatids. Compromised polarization of Par6 (**b**), Cdc42 (**d**) and PKC λ (**f**) in *Jam-C*^{-/-} round spermatids (RS KO). PATJ is localized in ES junctional plaques and a fraction is polarized in RS (**g**, **h**). Residual PATJ staining labels the *Jam-C*^{-/-} Sertoli-spermatid contact sites (**h**).

and nectin-3. Targeted inactivation of *nectin-2* in mice caused abnormal actin bundling at the Sertoli-spermatid junction, and male infertility owing to malformations of spermatozoan heads^{13–15}. We observed similar cytoskeletal defects in *Jam-C*^{-/-} spermatids, that is, F-actin was abnormally distributed and bundles failed to form at Sertoli-spermatid junctions (Fig. 2g–j). However, spermatid elongation was not blocked in *nectin-2* mutants so nectin and JAM molecules seem to have distinct functional roles. Likewise, targeted inactivation of the *hrb* gene (that is required for acrosome formation and male fertility), did not arrest polarization and formation of spermatid heads and flagella¹⁶. Thus, defects in cytoskeletal organization or acrosome formation were unlikely causes for the failure of *Jam-C*^{-/-} spermatids to polarize. To explore whether JAM-C has a direct role in cell polarization, we studied the cellular distribution of known regulators of this process. Partitioning-defective (Par) proteins regulate zygote polarity in response to sperm cues in the nematode *Caenorhabditis elegans*. Par6 and Par3 proteins can interact with the small GTPase Cdc42 and atypical protein kinase C (PKC) to form complexes mediating the establishment of cell polarity and polarized actin synthesis in many cell types and species^{17–20}. We found that Par6, Cdc42 and the serine/threonine kinase PKC λ show extensive co-localization with JAM-C in the heads of wild-type spermatids (Fig. 3a, c, e; and Supplementary Fig. 4). Whereas Par6, Cdc42 and PKC λ distribution was polarized in wild-type round spermatids, all three proteins showed little or no appreciable polarization in *Jam-C*^{-/-} spermatids (Fig. 3b, d, f). It therefore seems that JAM-C is necessary for the recruitment of a Par6–Cdc42–PKC λ complex to the anterior part of the prospective spermatid head and that this process is

crucial for spermatid differentiation.

Previous work has shown that the first PDZ domain of Par3 can bind to the carboxy terminus of JAM proteins and thereby recruit PKC λ to tight junctions in cultured cells^{21,22}. However, Par3 localization in spermatids showed only a little overlap with JAM-C and appeared largely unaffected in *Jam-C*^{-/-} spermatids (Supplementary Fig. 4). The multi-PDZ domain protein PATJ/Inadl is another regulator of tight junction assembly and cell polarity—this is through its interactions with the membrane-associated guanylate kinase PALS1 and a member of the crumbs (CRB) family, the transmembrane protein CRB3 (refs 23, 24). Recently identified interactions between Par6 and PALS1 as well as Par6 and PATJ have established direct connections between the two known macromolecular complexes mediating cell polarity, that is, Par6–Cdc42–aPKC and CRB–PALS1–PATJ^{25,26}. The spatial distribution of PATJ shows remarkable overlap with JAM-C, that is, a fraction is assembled in a polarized plaque whereas the rest is widely distributed in round spermatids (Fig. 3g, h). In the absence of JAM-C, the PATJ pattern was significantly altered but a fraction remained at the Sertoli cell-spermatid junction (Fig. 3h). It seems feasible that CRBs or other cell surface interaction partners may be involved in the control of PATJ cellular localization independently or even in competition with JAM-C. PATJ, through its ten PDZ domains, may also bind several partners simultaneously and facilitate the assembly of a large complex.

We next wanted to test whether engagement of JAM-C is sufficient for cell polarity protein recruitment. To mimic binding of JAM-C to other JAM proteins at the Sertoli-spermatid junction, soluble versions of JAM-B or JAM-C fused to the human immuno-

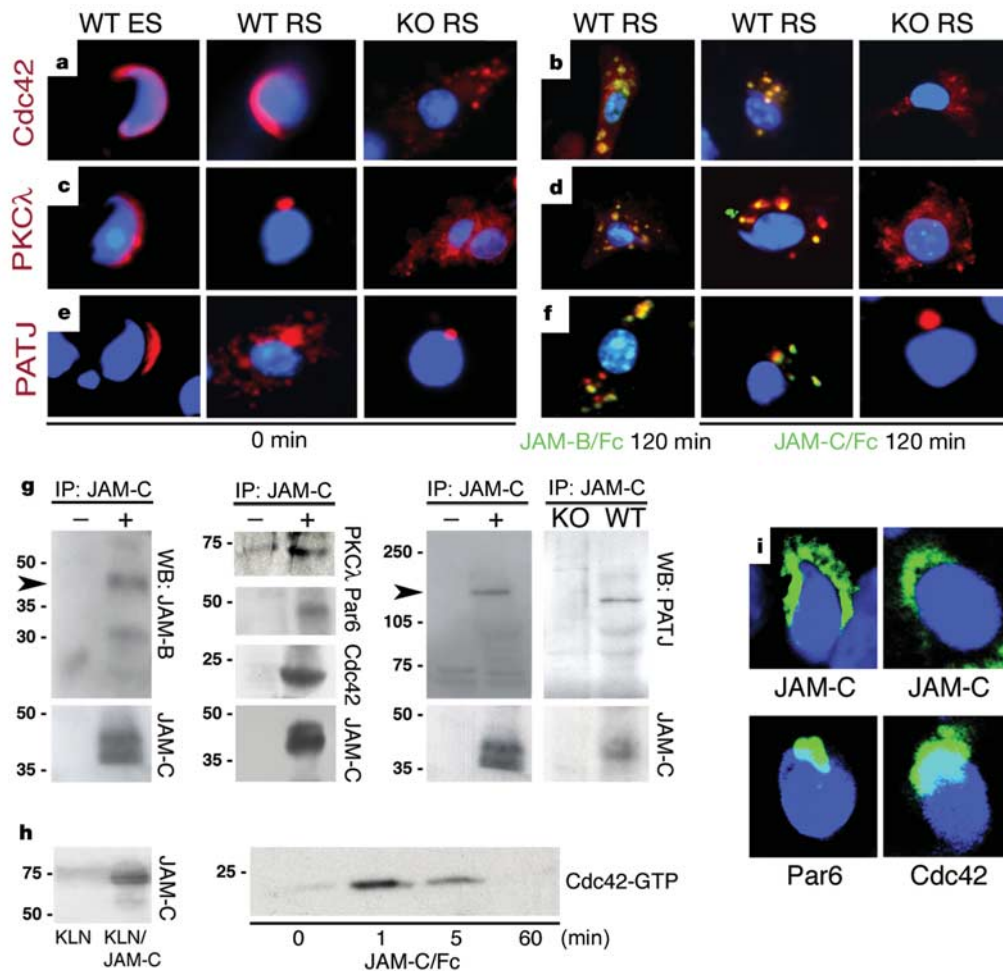


Figure 4 Cell polarity protein recruitment. **a, c, e**, Immunocytochemistry of isolated spermatids labelled with DAPI (blue) and antibodies against Cdc42 (**a**), PKC λ (**c**) or PATJ (**e**) (red). **b, d, f**, WT or KO spermatids were treated with clustered soluble JAM-B/Fc or JAM-C/Fc proteins (green label) and stained as indicated. Recruitment of Cdc42 (**b**), PKC λ (**d**) and PATJ (**f**) (red) to fusion protein binding sites (yellow) in WT spermatids. **g**, Association of JAM-B, Par6, Cdc42, PKC λ and PATJ with JAM-C in

immunoprecipitation (IP) from mouse testes. Absence (–) or presence (+) of the anti-JAM-C antibody is indicated. PATJ was absent when *Jam-C*^{–/–} testes (KO) were used for anti-JAM-C IP. Antibodies used in western blot (WB) are indicated. **h**, Stable expression of EGFP–JAM-C in KLN205 (KLN) cells. Stimulation with JAM-C/Fc leads to activation of Cdc42 (Cdc42-GTP) in KLN/JAM-C cells. **i**, Confocal images of human elongated and round spermatids stained for JAM-C, Par6 or Cdc42 (green).

globulin Fc portion (JAM-B/Fc, JAM-C/Fc) were applied to cultured primary spermatids. Before stimulation, the localization of Cdc42, PKC λ and a fraction of PATJ was polarized and resembled the patterns seen *in vivo*. Similarly, *Jam-C*^{–/–} spermatids showed wide distribution of Cdc42 and PKC λ , but polarized PATJ (Fig. 4a, c, e). Following incubation with JAM-B/Fc or JAM-C/Fc, Cdc42, PKC λ and PATJ were redistributed to the fusion protein binding sites in wild-type cells. Parallel stimulation experiments with spermatids from knockout (KO) mice demonstrated that expression of JAM-C is necessary for the binding of JAM-B/Fc and JAM-C/Fc, and, consequently, recruitment of Cdc42, PKC λ and PATJ (Fig. 4b, d, f; and data not shown). To support the relevance of the JAM–JAM interactions *in vivo*, we were able to co-immunoprecipitate complexes between JAM-B and JAM-C from testes. Likewise, Par6, Cdc42, PKC λ and PATJ were associated with immunoprecipitated JAM-C (Fig. 4g). Recent work in *Drosophila* embryos has shown that polarized localization of Par6 requires an interaction with activated Cdc42 (ref. 27). We found that stimulation of JAM-C-overexpressing cells with JAM-C/Fc fusion proteins (Fig. 4h) led to a rapid increase in the level of activated Cdc42, which in turn might lead to Par6 recruitment and cell polarity complex assembly.

To explore whether JAM-C and cell polarity proteins might have similar roles in human spermiogenesis, we studied the localization of JAM-C, Par6 and Cdc42 in sections of human testes. JAM-C was present in human spermatids and its distribution was already very polarized at the round cell stage. Likewise, Par6 and Cdc42 signals were highly polarized in patterns similar to what we had observed in mice (Fig. 4i). Therefore, it seems that the role of JAM-C and downstream cell polarity regulators may be conserved and relevant for human reproduction.

In conclusion, our results show that JAM-C is required for the polarization of round spermatids, which we attribute to its role in the recruitment of a cell polarity complex. We propose that the activity of Par6, Cdc42, PKC λ and PATJ downstream of engaged JAM-C protein mediates spermatid polarization in mice and possibly in humans. It seems feasible that genetic defects as well as other factors interfering with normal JAM-C function in the human reproductive system could lead to male infertility. □

Methods

Generation of *Jam-C* mutants

An 870 base pair (bp) *Jam-C* gene fragment containing exons 4 and 5 was flanked by loxP recognition sites for Cre recombinase. Within the loxP-flanked region, a murine *Jam-C*

cDNA fragment (codons 174–311, YRND...SFVI*) followed by the bovine growth hormone polyadenylation signal (pA) was fused to a *KpnI* site in exon 5. As an expression reporter, an IRES-lacZ-pA cassette was fused to the 5'-end of exon 4 and 260 bp of the upstream intron. The construct also contained a neomycin resistance cassette surrounded by *frt* sites enabling Flp recombinase-mediated removal, as well as long (8.8 kilobase (kb), 5') and short (1.3 kb, 3') arms for homologous recombination. Electroporated and G418-selected embryonic stem cell clones were analysed by Southern blot hybridization and polymerase chain reaction (PCR) (primers were 5'-TCATGTGGATAGCCACAACA-3' and 5'-ACATCTGTGCGACCTGCCA-3'; products were 350 bp from WT and 450 bp from targeted alleles). Two independent lines were generated and maintained in a mixed 129Sv $\times$ C57BL/6 background. JAM-C-deficient mutants were generated by cross-breeding with the PKG-Cre line²⁸ followed by interbreeding of *Jam-C* KO/+heterozygotes.

All animal experiments were performed in compliance with the relevant laws and were approved by the CR-UK LRI animal ethics committees.

Electron microscopy

For electron microscopic examination, testes were fixed in 2.5% glutaraldehyde buffered with Sorensen's phosphate (pH 7.4), and treated with 2% buffered OsO₄. Samples were dehydrated in a graded ethanol series and embedded in Araldite. Ultra-thin sections were examined by transmission electron microscopy using a JEOL 1010 microscope. Semi-thin sections were stained with Toluidine Blue in borax for light microscopic examination.

Immunofluorescence on tissue sections and lacZ staining

Cryosections (5 μ m) of testes in Cryomatrix (Thermo Shandon) were fixed with methanol or acetone for 10 min, incubated in blocking solution for 1 h (2% goat serum and 2% BSA in PBS), and incubated with primary antibodies diluted in blocking solution for 4 h or overnight. The following primary antibodies were used: rat monoclonal JAM-A H202 and JAM-C CRAM XIXH36; rabbit polyclonal JAM-B 829 (M.A.-L.); polyclonal Cdc42 and PKC α (Santa Cruz); polyclonal Par3 C2-3 (ref. 29); polyclonal Par6 (gift from P. Parker/T. Pawson); rabbit polyclonal PATJ²³; mouse monoclonal occludin and rabbit polyclonal ZO-1 (Zymed); and polyclonal Claudin-11 (Abcam). Appropriate species-specific secondary antibodies (Molecular Probes) were applied at 1:800 for 45 min. Phalloidin and peanut agglutinin were covalently linked to Alexa Fluor488 (Molecular Probes).

LacZ staining was performed as published previously³⁰.

In vitro recruitment assay

For the isolation of round spermatids, seminiferous tubules from decapsulated testes were minced and suspended in PBS. Following centrifugation (200 g for 1 min), the supernatant was collected and the pellet was resuspended in PBS. The centrifugation procedure was repeated three times. Supernatants from the spinning cycles were pooled and applied onto a 10 μ m cell micro sieve (VWR International) to enrich for round spermatids. Cells were collected by centrifugation at 1,000 g for 15 min, resuspended in MEM/Ham's F12 medium supplemented with 1 mM sodium pyruvate, 15 mM HEPES, 6 mM D-L-lactic acid, 10% FCS and antibiotics, plated on poly-L-lysine coated coverslips, and incubated overnight at 37°C.

For recruitment studies, cells were treated for 2 h with serum-free medium containing 2 μ g ml⁻¹ JAM-C/Fc or JAM-B/Fc chimera (R&D Systems) pre-clustered with goat anti-human IgG Fc (Jackson Laboratories). Cells were then washed, fixed with methanol for 3 min, blocked with 2% goat serum, 2% horse serum and 2% BSA in PBS for 1 h, and incubated with primary antibodies for 3 h. Appropriate species-specific secondary antibodies were then applied in combination with donkey anti-goat Alexa Fluor488 for visualization of bound JAM fusion proteins.

Immunoprecipitations

Testes were isolated, decapsulated and homogenized in 10 mM Tris-HCl (pH 7.5), 0.32 M sucrose, 3 mM MgCl₂, 10 mM iodoacetamide, 0.5% aprotinin and 0.5% leupeptin, using a Polytron homogenizer. The supernatant of a first centrifugation (1,000 g, 30 min) was spun at 100,000 g for 60 min at 4°C. The pellet was solubilized in TNT buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl and 0.5% Triton X-100) in the presence of protease inhibitors. Pre-cleared lysates were incubated with antibody-coupled protein G beads overnight at 4°C. Following washes with TNT buffer, beads were resuspended in Laemmli buffer and proteins were analysed by SDS–polyacrylamide gel electrophoresis (PAGE) and western blotting.

Cdc42 activation assay

An activation assay kit (Upstate) was used to precipitate Cdc42-GTP from KLN205 cells (CRL1453, ATCC) stably overexpressing JAM-C, which was extracellularly tagged by insertion of enhanced green fluorescent protein (EGFP) between the C2 domain and the transmembrane region (M.A.-L. unpublished data). Cells were plated in 90 mm dishes and grown to 80–90% confluence, serum-starved (12 h), treated for the indicated times with serum-free medium containing 2 μ g ml⁻¹ pre-clustered JAM-C/Fc chimera (R&D Systems), washed with PBS and lysed with ice-cold MLB (25 mM HEPES, pH 7.5, 150 mM NaCl, 1% Igepal CA-630, 10 mM MgCl₂, 1 mM EDTA, 10% glycerol and protease inhibitors). Lysates were spun for 5 min at 14,000 g (4°C) and supernatants were incubated with 5 μ g of PAK1 PBD agarose for 45 min at 4°C. Agarose beads were washed three times with MLB and resuspended in 40 μ l 2 $\times$ Laemmli sample buffer. Following SDS–PAGE and western blot, Cdc42 was detected with a mouse monoclonal antibody (Upstate).

Received 20 June; accepted 19 July 2004; doi:10.1038/nature02877.

- Cheng, C. Y. & Mruk, D. D. Cell junction dynamics in the testis: Sertoli-germ cell interactions and male contraceptive development. *Physiol. Rev.* **82**, 825–874 (2002).
- Griswold, M. D. Interactions between germ cells and Sertoli cells in the testis. *Biol. Reprod.* **52**, 211–216 (1995).
- Ezesh, U. I. Beyond the clinical classification of azoospermia: opinion. *Hum. Reprod.* **15**, 2356–2359 (2000).
- Muller, W. A. Leukocyte–endothelial-cell interactions in leukocyte transmigration and the inflammatory response. *Trends Immunol.* **24**, 327–334 (2003).
- Aurrand-Lions, M., Johnson-Leger, C. & Imhof, B. A. Role of interendothelial adhesion molecules in the control of vascular functions. *Vascul. Pharmacol.* **39**, 239–246 (2002).
- Chavakis, T., Preissner, K. T. & Santoso, S. Leukocyte trans-endothelial migration: JAMs add new pieces to the puzzle. *Thromb. Haemost.* **89**, 13–17 (2003).
- Bazzoni, G. The JAM family of junctional adhesion molecules. *Curr. Opin. Cell Biol.* **15**, 525–530 (2003).
- Ebnet, K., Suzuki, A., Ohno, S. & Vestweber, D. Junctional adhesion molecules (JAMs): more molecules with dual functions? *J. Cell Sci.* **117**, 19–29 (2004).
- Liang, T. W. *et al.* Vascular endothelial-junctional adhesion molecule (VE-JAM)/JAM 2 interacts with T, NK, and dendritic cells through JAM 3. *J. Immunol.* **168**, 1618–1626 (2002).
- Arrate, M. P., Rodriguez, J. M., Tran, T. M., Brock, T. A. & Cunningham, S. A. Cloning of human junctional adhesion molecule 3 (JAM3) and its identification as the JAM2 counter-receptor. *J. Biol. Chem.* **276**, 45826–45832 (2001).
- Lui, W. Y., Mruk, D., Lee, W. M. & Cheng, C. Y. Sertoli cell tight junction dynamics: their regulation during spermatogenesis. *Biol. Reprod.* **68**, 1087–1097 (2003).
- Kierszenbaum, A. L., Rivkin, E. & Tres, L. L. Acroplaxome, an F-actin-keratin-containing plate, anchors the acrosome to the nucleus during shaping of the spermatid head. *Mol. Biol. Cell* **14**, 4628–4640 (2003).
- Ozaki-Kuroda, K. *et al.* Nectin couples cell–cell adhesion and the actin scaffold at heterotypic testicular junctions. *Curr. Biol.* **12**, 1145–1150 (2002).
- Bouchard, M. J. *et al.* Defects in nuclear and cytoskeletal morphology and mitochondrial localization in spermatozoa of mice lacking nectin-2, a component of cell–cell adherens junctions. *Mol. Cell. Biol.* **20**, 2865–2873 (2000).
- Mueller, S., Rosenquist, T. A., Takai, Y., Bronson, R. A. & Wimmer, E. Loss of nectin-2 at Sertoli-spermatid junctions leads to male infertility and correlates with severe spermatozoan head and midpiece malformation, impaired binding to the zona pellucida, and oocyte penetration. *Biol. Reprod.* **69**, 1330–1340 (2003).
- Kang-Decker, N., Mantchev, G. T., Juneja, S. C., McNiven, M. A. & van Deursen, J. M. Lack of acrosome formation in Hrb-deficient mice. *Science* **294**, 1531–1533 (2001).
- Etienne-Manneville, S. & Hall, A. Cell polarity: Par6, aPKC and cytoskeletal crosstalk. *Curr. Opin. Cell Biol.* **15**, 67–72 (2003).
- Ohno, S. Intercellular junctions and cellular polarity: the PAR-aPKC complex, a conserved core cassette playing fundamental roles in cell polarity. *Curr. Opin. Cell Biol.* **13**, 641–648 (2001).
- Erickson, J. W. & Cerione, R. A. Multiple roles for Cdc42 in cell regulation. *Curr. Opin. Cell Biol.* **13**, 153–157 (2001).
- Wodarz, A. Establishing cell polarity in development. *Nature Cell Biol.* **4**, E39–E44 (2002).
- Ebnet, K. *et al.* The junctional adhesion molecule (JAM) family members JAM-2 and JAM-3 associate with the cell polarity protein PAR-3: a possible role for JAMs in endothelial cell polarity. *J. Cell Sci.* **116**, 3879–3891 (2003).
- Ebnet, K. *et al.* The cell polarity protein ASIP/Par-3 directly associates with junctional adhesion molecule (JAM). *EMBO J.* **20**, 3738–3748 (2001).
- Roh, M. H., Fan, S., Liu, C. J. & Margolis, B. The Crumbs3–Pals1 complex participates in the establishment of polarity in mammalian epithelial cells. *J. Cell Sci.* **116**, 2895–2906 (2003).
- Straight, S. W. *et al.* Loss of PALS1 expression leads to tight junction and polarity defects. *Mol. Biol. Cell* **15**, 1981–1990 (2004).
- Nam, S. C. & Choi, K. W. Interaction of Par-6 and Crumbs complexes is essential for photoreceptor morphogenesis in *Drosophila*. *Development* **130**, 4363–4372 (2003).
- Hurd, T. W., Gao, L., Roh, M. H., Macara, I. G. & Margolis, B. Direct interaction of two polarity complexes implicated in epithelial tight junction assembly. *Nature Cell Biol.* **5**, 137–142 (2003).
- Hutterer, A., Betschinger, J., Petronczki, M. & Knoblich, J. A. Sequential roles of Cdc42, Par-6, aPKC and Lgl in the establishment of epithelial polarity during *Drosophila* embryogenesis. *Dev. Cell* **6**, 845–854 (2004).
- Lallemand, Y., Luria, V., Haffner-Krausz, R. & Lonai, P. Maternally expressed PKG-Cre transgene as a tool for early and uniform activation of the Cre site-specific recombinase. *Transgenic Res.* **7**, 105–112 (1998).
- Izumi, Y. *et al.* An atypical PKC directly associates and colocalizes at the epithelial tight junction with ASIP, a mammalian homologue of *Caenorhabditis elegans* polarity protein PAR-3. *J. Cell Biol.* **143**, 95–106 (1998).
- Compagni, A., Logan, M., Klein, R. & Adams, R. H. Control of skeletal patterning by ephrinB1–EphB interactions. *Dev. Cell* **5**, 217–230 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank B. Margolis, T. Pawson, E. Peles, A. Suzuki, A. Le Bivic, C. Nobes and P. Parker for reagents and N. Hogg for comments on the manuscript. This work was funded by Cancer Research UK, the EMBO Young Investigator Program (R.H.A.), the Swiss National Science Foundation (M.A.L. and B.A.L.) and Oncosuisse (B.A.L.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.H.A. (ralf.adams@cancer.org.uk).

A microtubule-binding myosin required for nuclear anchoring and spindle assembly

Kari L. Weber^{1,*}, Anna M. Sokac^{1,2,*}, Jonathan S. Berg³, Richard E. Cheney³ & William M. Bement^{1,2}

¹Department of Zoology, ²Program in Cellular and Molecular Biology, University of Wisconsin, Madison, Madison, Wisconsin 53706, USA

³Department of Cell and Molecular Physiology, University of North Carolina, Chapel Hill, Chapel Hill, North Carolina 27599, USA

* Present addresses: Department of Cell Biology, CB163, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, California 92037, USA (K.L.W.); Department of Molecular Biology, Princeton University, Washington Road, Princeton, New Jersey 08544, USA (A.M.S.)

Proper spindle positioning and orientation are essential for asymmetric cell division and require microtubule–actin filament (F-actin) interactions in many systems^{1,2}. Such interactions are particularly important in meiosis³, where they mediate nuclear anchoring^{4–6}, as well as meiotic spindle assembly and rotation^{7,8}, two processes required for asymmetric cell division. Myosin-10 proteins are phosphoinositide-binding⁹, actin-based motors that contain carboxy-terminal MyTH4 and FERM domains of unknown function¹⁰. Here we show that *Xenopus laevis* myosin-10 (Myo10) associates with microtubules *in vitro* and *in vivo*,

and is concentrated at the point where the meiotic spindle contacts the F-actin-rich cortex. Microtubule association is mediated by the MyTH4-FERM domains, which bind directly to purified microtubules. Disruption of Myo10 function disrupts nuclear anchoring, spindle assembly and spindle–F-actin association. Thus, this myosin has a novel and critically important role during meiosis in integrating the F-actin and microtubule cytoskeletons.

Microtubule–F-actin interactions are essential for a broad variety of basic biological processes that require polarized distribution of cellular components². Mitotic spindle positioning in yeast, for example, is critically dependent on physical linkages between microtubules and cortical F-actin¹. Microtubule–F-actin interactions are particularly important in oogenesis, which entails anchoring and translocation of microtubule-based structures within a comparatively large cell volume³. For example, the oocyte nucleus (germinal vesicle) is anchored by microtubules that extend from the nucleus to the F-actin-rich cortex^{4–6}, whereas proper meiotic spindle assembly and/or targeting to the plasma membrane are F-actin dependent^{3,7,8,11}. Further, the site of asymmetric spindle anchoring at the oocyte cortex is overlaid by a cap of high F-actin density in mammals^{8,11}, amphibians⁷ and invertebrates^{12,13}; this cap is thought to anchor the spindle at the cortex through interaction with microtubules^{3,7,8,11}. At present, however, the molecular mechanisms that couple microtubules to F-actin during oogenesis are unknown.

Xenopus Myo10 was identified in a polymerase chain reaction

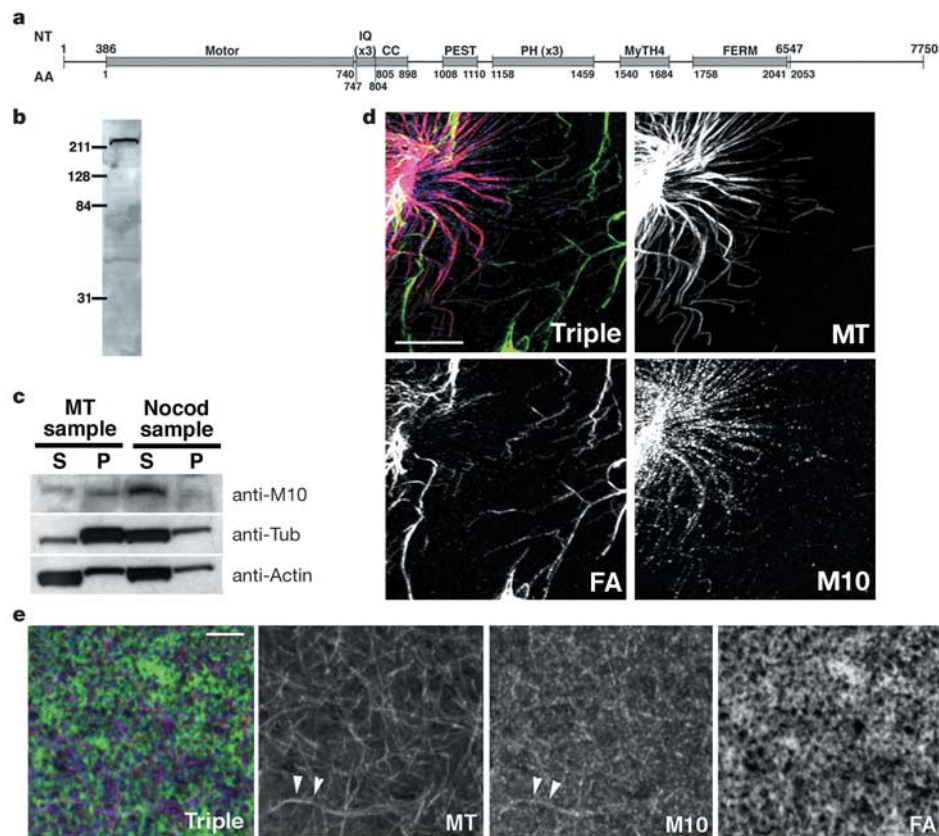


Figure 1 Myo10 associates with microtubules *in vitro* and *in vivo*. **a**, Domain organization of Myo10. **b**, Myo10 antibody recognizes single protein of ~230 kDa in egg extract immunoblot. **c**, Microtubule (P, pellet) and supernatant fractions (S) from egg extracts induced to (MT sample) or prevented from (Nocod sample) polymerizing microtubules (see Methods). Immunoblotting with antibodies to Myo10 (M10), tubulin (Tub) or actin

(Actin) shows Myo10 cofractionates with microtubules but not actin filaments. **d**, Myo10 (M10, blue) immunolocalizes with aster microtubules (MT, red) but much less with F-actin (FA, green) in extracts. Scale bar, 30 μ m. **e**, Myo10 also colocalizes with microtubules in intact eggs (arrowheads). Antibodies the same as in **d**. Scale bar, 20 μ m.

(PCR) screen for myosins expressed in oocytes. Although myosins are best known as F-actin-based motors¹⁴, we characterized Myo10 as a potential microtubule–F-actin linker in oocytes, based on its primary structure. That is, this unconventional vertebrate myosin contains a MyTH4 domain in its C-terminal tail¹⁰, and the MyTH4 domain of an *Arabidopsis* kinesin-like protein was previously shown to bind microtubules¹⁵. The complete coding sequence of Myo10 was obtained (Fig. 1a; Supplementary Fig. S1) and polyclonal antibodies (directed against a small region in the actin-based motor ‘head’ region) were generated and affinity purified (Fig. 1b). Consistent with potential microtubule-binding, Myo10 co-sedimented with microtubules (Fig. 1c) and colocalized with microtubules in *Xenopus* egg extracts (Fig. 1d) and in intact *Xenopus* eggs (Fig. 1e). The observed immunolocalization of Myo10 on microtubules was specific and independent of F-actin (Supplementary Fig. S2), as expected if Myo10 bound directly to microtubules rather than associating indirectly through actin binding.

Myo10 showed a striking colocalization with meiotic spindle microtubules (Fig. 2). Specifically, Myo10 concentrated in the region where the spindle contacted the cortex (Fig. 2a, b), and on microtubules that extended upward from the spindle into the cortex, similar to guy wires (Fig. 2c, d). F-actin also colocalized with the spindle, both in the cortical cap and the interior of the spindle. Although some overlap of Myo10 and F-actin staining was observed, Myo10 was most abundant in the outer regions of the spindle and the cap, whereas F-actin was most abundant on the interior regions of the cap and the spindle (Fig. 2).

To determine the minimal region of Myo10 sufficient for microtubule association, a series of enhanced green fluorescent protein (eGFP)–Myo10 truncation constructs were expressed in oocytes by messenger RNA injection (Fig. 3a). Co-sedimentation analysis showed that the MyTH4-FERM domain cassette was sufficient to

confer microtubule association (Fig. 3b). The association was independent of F-actin, as demonstrated by pretreating oocytes with latrunculin (Supplementary Fig. S3). To further dissect the region of Myo10 required for microtubule association and to determine whether microtubule binding was direct, we sought to express Myo10 tail constructs in bacteria. However, these efforts were frustrated by insolubility and instability of the expressed constructs. As an alternative approach, reticulocyte lysate was programmed with mRNA encoding eGFP alone or eGFP fused to MyTH4, FERM or the MyTH4-FERM cassette. After expression, lysate was combined with pure microtubules and analysed by co-sedimentation. Neither eGFP alone nor eGFP–FERM pelleted with microtubules, whereas eGFP–MyTH4 showed modest co-pelleting and eGFP–MyTH4-FERM showed significant co-pelleting (Fig. 3c), demonstrating that the MyTH4-FERM cassette was sufficient for microtubule association.

Because it could be argued that a component of the reticulocyte lysate was responsible for linking the MyTH4-FERM cassette to microtubules, mRNA encoding glutathione *S*-transferase (GST) alone or GST–MyTH4-FERM was injected into oocytes, and after translation, purified through glutathione affinity (Supplementary Fig. S4). GST–MyTH4-FERM pelleted in the presence, but not the absence of microtubules, whereas GST alone did not pellet in the presence of microtubules (Fig. 3d). Thus, the MyTH4-FERM domain cassette of Myo10 binds directly to microtubules.

To assess the *in vivo* role of Myo10, oocytes were injected with mRNA encoding truncated proteins consisting of the pleckstrin homology (PH), MyTH4 and FERM domains fused to either eGFP or GST. Tail truncates function as dominant negatives for unconventional myosins in general¹⁶ and more specifically, for myosin-10 in phagocytosis¹⁷. Both eGFP–Myo10 tail (Fig. 4a, b) and GST–Myo10 tail (data not shown) resulted in displacement of the oocyte

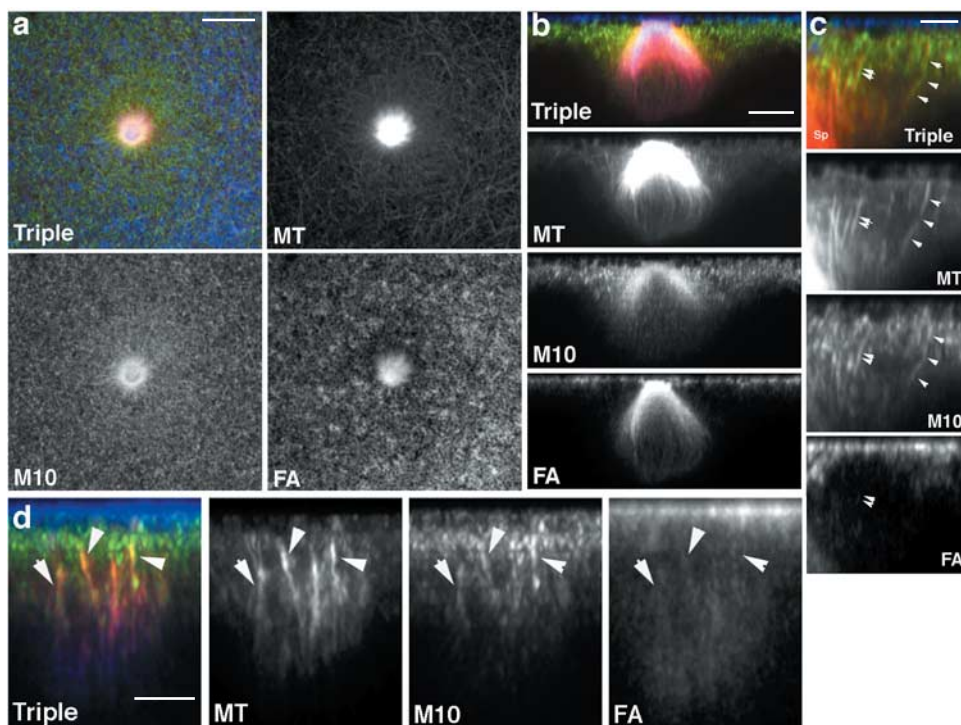


Figure 2 Myo10 localizes to meiotic spindles. **a**, Surface view shows colocalization of Myo10 (M10, green) and F-actin (FA, blue) with meiotic spindle (MT, red). Scale bar, 30 μ m. **b**, Z-section shows that Myo10 and F-actin have overlapping but distinct distributions on the spindle with Myo10 predominantly around the outside and F-actin on the inside of the spindle. Scale bar, 15 μ m. **c**, High magnification Z-section shows tracks

of Myo10 on microtubules that extend upward from the spindle (sp) to the cortex (arrowheads). Tracks both with (double arrowheads) and without (single arrowheads) associated F-actin are seen. Scale bar, 5 μ m. **d**, Grazing view of spindle shows Myo10 concentrated on spindle microtubules extending into the cortex (arrowheads). Scale bar, 5 μ m.

nucleus (germinal vesicle) from its normal site of asymmetric localization in the animal hemisphere to the cortex, as shown by a large white spot at the animal pole. Consistent with results from other systems¹⁷, displacement required the entire tail. The specificity of this effect was demonstrated by the fact that co-expression of eGFP–full length Myo10, but not coexpression of eGFP alone, suppressed nuclear displacement by the tail construct (Fig. 4b).

The oocyte nucleus is asymmetrically positioned by microtubules that extend from its surface to the F-actin rich cortical cytoskeleton^{5,18}. Consequently, microtubule depolymerization in both amphibian and mammalian oocytes results in nuclear displacement, as the nucleus is less dense than the surrounding cytoplasm^{5,6} (Fig. 4a). Thus, it was possible that disruption of Myo10 function resulted in microtubule depolymerization. However, although expression of the dominant negative Myo10 tail mimicked the phenotype produced by microtubule depolymerization (Fig. 4a), they did so without causing microtubule depolymerization, as judged by both immunofluorescence analysis and biochemical quantification of polymeric tubulin (Supplementary Fig. S5).

As an independent means to test the *in vivo* role of Myo10, oocytes were microinjected with the affinity-purified Myo10 antibody. Anti-Myo10 microinjection resulted in the same phenotype as the dominant negative Myo10 tail expression, causing germinal vesicle displacement (Fig. 4c). This effect was both dose-dependent and specific, in that microinjection of a non-specific IgG failed to result in germinal vesicle displacement (Fig. 4c). Thus, Myo-10 is

required for microtubule-dependent, asymmetric anchoring of the germinal vesicle.

Meiotic spindle assembly and/or positioning are F-actin dependent in oocytes^{3,7,8,11}. To assess the potential role of Myo10 in spindle assembly, oocytes injected with either dominant negative Myo10 tail or anti-Myo10 antibodies were induced to undergo meiotic maturation by treatment with progesterone. The dominant negative Myo10 tail construct targeted to spindles (Supplementary Fig. S6) and the plasma membrane (data not shown) and this construct, as well as microinjection of Myo10 antibodies, severely impaired meiotic spindle assembly (Fig. 4d–f). The phenotypes produced, ranging from rotation failure and abnormal spindle structure, to multiple microtubule organizing centres (Figs 4d; Supplementary Fig. S7), mimic those previously described that result from actin depolymerization during meiotic maturation⁷. However, neither the Myo10 tail nor microinjection of Myo10 antibody resulted in actin disassembly. Rather, these manipulations resulted in abnormal F-actin–spindle association such that F-actin concentrated in aggregates on or near spindles or abortive microtubule organizing centres (Fig. 4d), indicating that Myo10 is essential for proper F-actin–meiotic spindle association. Thus, Myo10 disruption produced phenotypes characteristic of microtubule disassembly (nuclear displacement) or F-actin disassembly (failed spindle assembly) without disassembling either polymer system, indicating that this actin-based motor serves to link the microtubule and cortical F-actin cytoskeletons.

The identification of Myo10 as an essential integrator of micro-

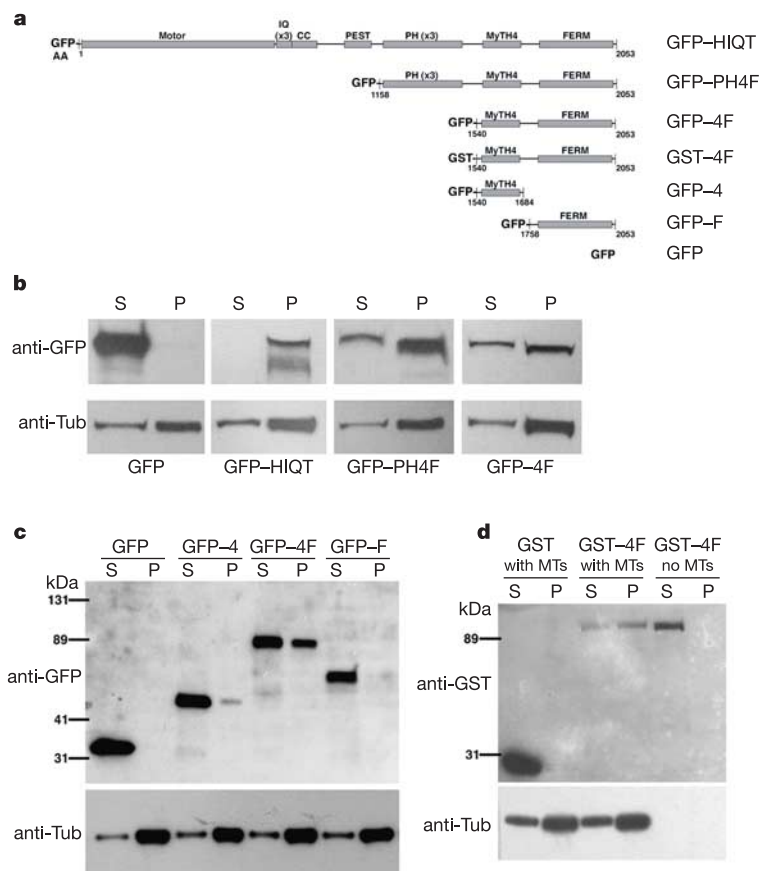


Figure 3 The MyTH4-FERM domain cassette of Myo10 binds microtubules. **a**, Constructs used for experiments in Figs 3 and 4. **b**, Immunoblot of the co-sedimentation assay shows that constructs containing the MyTH4 and FERM domains expressed through mRNA injection co-pellet with oocyte microtubules. **c**, Immunoblot of the co-sedimentation assay

showing that both MyTH4 and FERM are required for optimal microtubule association in reticulocyte lysate. **d**, Immunoblot showing that GST–MyTH4-FERM but not GST purified following oocyte expression bind purified microtubules.

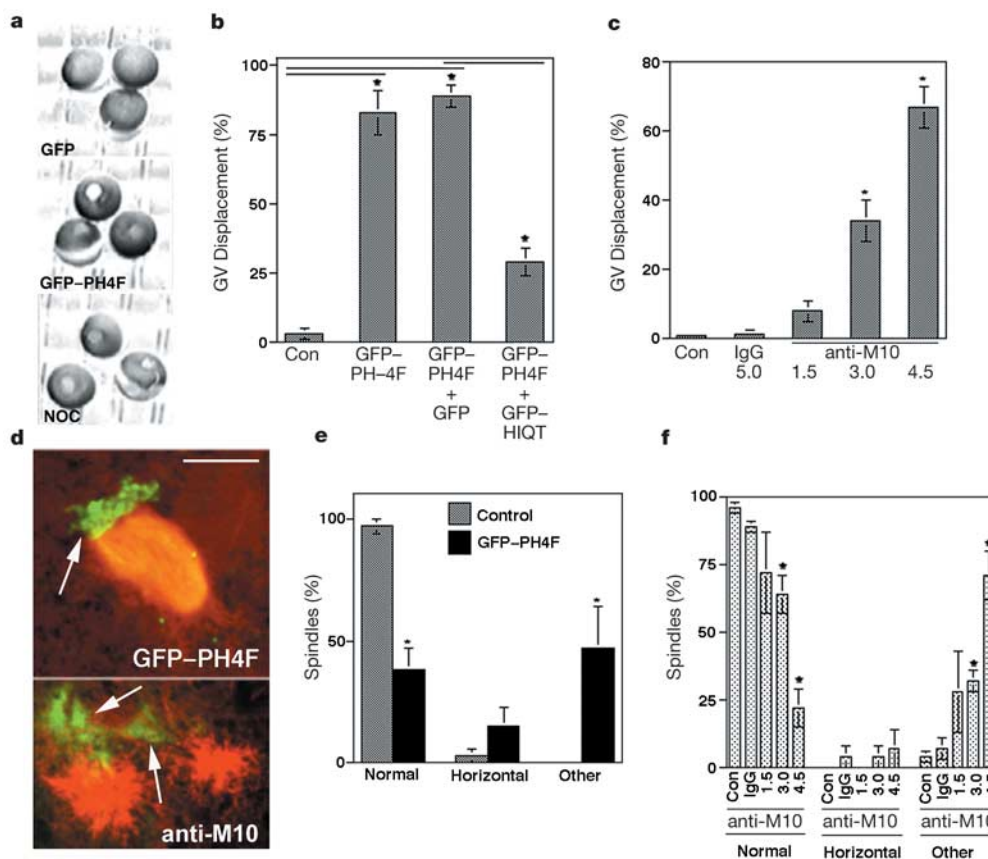


Figure 4 Myo10 is required for nuclear anchoring and meiotic spindle assembly. **a**, GFP-PH4F expression results in nuclear (germinal vesicle, GV) displacement in oocytes (shown by the white spot) not seen in controls that expressed GFP alone. Nocodazole (NOC) produced the same phenotype. **b**, The germinal vesicle displacement phenotype is suppressed by co-expression of GFP-HIQT but not GFP alone. **c**, Microinjection of the Myo10 antibody (M10) at increasing concentrations (numbers indicate concentration in needle, mg ml^{-1}) also resulted in germinal vesicle displacement. Non-specific IgG (IgG) had no effect. Lines above the bars indicate groups compared for statistical analysis. **d**, In

egg expressing dominant negative Myo10 (cGFP-PH4F, upper panel), F-actin (pseudocoloured green) is found in aggregate (arrow) at one end of an unrotated, malformed spindle (red). In egg microinjected with anti-Myo10 antibody (M10, lower panel), two aberrant microtubule arrays (red) are present; F-actin (green) is in aggregates (arrows). Scale bar, 15 μm . **e**, GFP-PH4F expression impairs meiotic spindle assembly. **f**, Myo10 antibody also significantly impairs spindle assembly. Results are mean $\pm$ s.e.m. Asterisk indicates $P < 0.05$.

tubules with cortical F-actin in nuclear anchoring and spindle assembly in oocytes fits well with its known properties. That is, the PH domain-dependent binding of phosphoinositides^{9,19}, actin-binding and motor activity²⁰, and microtubule binding (this study) could function both passively and actively: passively for nuclear anchoring, through a stable association of nuclear microtubules with the cortex, where both phosphoinositides and F-actin are abundant; actively for spindle assembly, wherein the motor activity could be harnessed to help transport microtubule-associated spindle components over large distances. In addition, given the known association of myosin-10 with dynamic actin in other systems^{10,17}, an intriguing possibility is that the meiotic spindle cap may be composed of rapidly assembling actin and that myosin-10 could directly contribute to its assembly, through its association with Mena/VASP²¹. It will therefore be of great interest to determine whether association of myosin-10 with its various binding partners is differentially regulated, and whether such associations impact its motor activity.

This report has a further, critical implication. Namely, the demonstration that the Myo10 MyTH4-FERM domain cassette binds microtubules directly suggests that this cassette in other unconventional myosins may serve the same function. Of particular interest is myosin-7a. Genetic defects in myosin-7a result in Usher syndrome in humans²², and myosin-7a localizes to cilia in a variety of tissues²³. Because cilia are microtubule-rich, we suggest that this localization is mediated through the MyTH4-FERM domain of

myosin-7a. More generally, our results, as well as demonstrations showing that other unconventional myosins can associate with microtubules indirectly^{24,25} or directly²⁶ suggest that these motors may generally serve to integrate the F-actin and microtubule cytoskeletons. □

Methods

Oocyte culture, injection and immunofluorescence

Oocytes were obtained, microinjected and cultured as previously described²⁷. For microtubule depolymerization, oocytes were incubated in culture medium containing 20 μM nocodazole for 24 h. Oocytes were analysed 24 h after antibody injection or 48 h after mRNA injection. For immunofluorescence, eggs were fixed and stained for F-actin and microtubules as previously described²⁸. For extract experiments, immunofluorescence following rapid freezing was performed as described²⁹. In both cases Myo10 antibody was used at 3 $\mu\text{g ml}^{-1}$.

Construct generation and expression

A small fragment of Myo10 was amplified by degenerate PCR³⁰ and full length sequence was obtained using 5' and 3' rapid amplification of cDNA ends (RACE). RNA was made as previously described using a pCS2-eGFP vector²⁷. Oocytes were microinjected with 40 nl RNA at 1–4 mg ml^{-1} (needle concentration).

Antibody production and fractionation

Polyclonal antibodies against a fusion protein composed of GST and amino acids 269–390 of Myo10 were raised and affinity purified (after preclearing for GST-reactive antibodies). Prior to microinjection, antibodies were concentrated using microcon columns. For extracts, microtubule polymerization was induced or inhibited by addition of dimethylsulphoxide (DMSO) (20%) or nocodazole (20 μM), respectively, before pelleting and immunoblotting.

For reticulocyte lysate (Ambion) the manufacturer's instructions were followed for translation of added mRNA. Lysates were subjected to a 100,000g centrifugation to pellet insoluble material, and then the supernatants were combined with microtubules (assembled from purified tubulin) and pelleted by centrifugation at 100,000g. For expression and purification of GST and GST–MyTH4–FERM in oocytes, two days after injection, 75 oocytes were homogenized in PBS supplemented with protease inhibitors and 1% triton X-100. After a 5,000g clearing centrifugation, supernatants were mixed with glutathione–sepharose (Amersham) and incubated at 4 °C for 1 h. Glutathione–sepharose was then extracted 3 × with 1 M NaCl in PBS to remove any proteins associated with the constructs, and eluted with 10 mM glutathione. Eluates were pooled and concentrated with microcon columns, and then combined with microtubules assembled from purified tubulin. Microtubule fractions were then pelleted by a 100,000g centrifugation.

Received 6 March; accepted 12 July 2004; doi:10.1038/nature02834.

1. Gundersen, G. G. & Bretscher, A. Cell biology. Microtubule asymmetry. *Science* **300**, 2040–2041 (2003).
2. Rodriguez, O. C. *et al.* Conserved microtubule–actin interactions in cell movement and morphogenesis. *Nature Cell Biol.* **5**, 599–609 (2003).
3. Sardet, C., Prodon, F., Dumollard, R., Chang, P. & Chenevert, J. Structure and function of the egg cortex from oogenesis through fertilization. *Dev. Biol.* **241**, 1–23 (2002).
4. Lessman, C. A. Germinal vesicle migration and dissolution in *Rana pipiens* oocytes: effect of steroids and microtubule poisons. *Cell Differ.* **20**, 239–251 (1987).
5. Gard, D. L. Ectopic spindle assembly during maturation of *Xenopus* oocytes: evidence for functional polarization of the oocyte cortex. *Dev. Biol.* **159**, 298–310 (1993).
6. Alexandre, H., Van Cauwenberge, A. & Mulnard, J. Involvement of microtubules and microfilaments in the control of the nuclear movement during maturation of mouse oocyte. *Dev. Biol.* **136**, 311–320 (1989).
7. Gard, D. L., Cha, B. J. & Roeder, A. D. F-actin is required for spindle anchoring and rotation in *Xenopus* oocytes: a re-examination of the effects of cytochalasin B on oocyte maturation. *Zygote* **3**, 17–26 (1995).
8. Sun, Q. Y. *et al.* Dynamic events are differently mediated by microfilaments, microtubules, and mitogen-activated protein kinase during porcine oocyte maturation and fertilization *in vitro*. *Biol. Reprod.* **64**, 879–889 (2001).
9. Isakoff, S. J. *et al.* Identification and analysis of PH domain-containing targets of phosphatidylinositol 3-kinase using a novel *in vivo* assay in yeast. *EMBO J.* **17**, 5374–5387 (1998).
10. Berg, J. S., Derfler, B. H., Pennisi, C. M., Corey, D. P. & Cheney, R. E. Myosin-X, a novel myosin with pleckstrin homology domains, associates with regions of dynamic actin. *J. Cell Sci.* **113**, 3439–3451 (2000).
11. Kim, N.-H. *et al.* The distribution and requirements of microtubules and microfilaments in bovine oocytes during *in vitro* maturation. *Zygote* **8**, 25–32 (2000).
12. Shimizu, T. Polar body formation in *Tubifex* eggs. *Ann. NY Acad. Sci.* **582**, 260–272 (1990).
13. Sardet, C., Speksnijder, J., Terasaki, M. & Chang, P. Polarity of the ascidian egg cortex before fertilization. *Development* **115**, 221–237 (1992).
14. Sokac, A. M. & Bement, W. M. Regulation and expression of metazoan unconventional myosins. *Int. Rev. Cytol.* **200**, 197–304 (2000).
15. Narasimhulu, S. B. & Reddy, A. S. N. Characterization of microtubule binding domains in the *Arabidopsis* kinesin-like calmodulin binding protein. *Plant Cell* **10**, 957–965 (1998).
16. Rogers, S. L. *et al.* Regulation of melanosome movement in the cell cycle by reversible association with myosin V. *J. Cell Biol.* **146**, 1265–1276 (1999).
17. Cox, D. *et al.* Myosin-X is a downstream effector of PI(3)K during phagocytosis. *Nature Cell Biol.* **4**, 469–477 (2002).
18. Gard, D. L. Organization, nucleation, and acetylation of microtubules in *Xenopus laevis* oocytes: a study by confocal immunofluorescence microscopy. *Dev. Biol.* **143**, 346–362 (1991).
19. Yonezawa, S. *et al.* Possible involvement of myosin-X in intercellular adhesion: importance of serial pleckstrin homology regions for intracellular localization. *Dev. Growth Differ.* **45**, 175–185 (2003).
20. Homma, K., Saito, J., Ikebe, R. & Ikebe, M. Motor function and regulation of myosin X. *J. Biol. Chem.* **276**, 34348–34354 (2001).
21. Tokuo, H. & Ikebe, M. Myosin-X transports Mena/VASP to the tip of filopodia. *Biochem. Biophys. Res. Commun.* **319**, 214–220 (2004).
22. Weil, D. *et al.* Defective myosin VIIA gene responsible for Usher syndrome type 1B. *Nature* **374**, 60–61 (1995).
23. Wolfrum, U., Liu, X., Schmitt, A., Udovichenko, I. P. & Williams, D. S. Myosin VIIa as a common component of cilia and microvilli. *Cell. Motil. Cytoskeleton* **40**, 261–271 (1998).
24. Lantz, V. A. & Miller, K. G. A class VI unconventional myosin is associated with a homologue of a microtubule-binding protein, cytoplasmic linker protein-170, in neurons and at the posterior pole of *Drosophila* embryos. *J. Cell Biol.* **140**, 897–910 (1998).
25. Huang, J. D. *et al.* Direct interaction of microtubule- and actin-based transport motors. *Nature* **397**, 267–270 (1999).
26. Cao, T. T., Chang, W., Masters, S. E. & Mooseker, M. S. Myosin-Va binds to and mechanochemically couples microtubules to actin filaments. *Mol. Biol. Cell* **15**, 151–161 (2004).
27. Sokac, A. M., Co, C., Taunton, J. & Bement, W. Cdc42-dependent actin polymerization during compensatory endocytosis in *Xenopus* eggs. *Nature Cell Biol.* **5**, 727–732 (2003).
28. Mandato, C. A. & Bement, W. M. Actomyosin transports microtubules and microtubules control actomyosin recruitment during *Xenopus* oocyte wound healing. *Curr. Biol.* **13**, 1096–1105 (2003).
29. Weber, K. L. & Bement, W. M. F-actin serves as a template for cytokeratin organization in cell free extracts. *J. Cell Sci.* **115**, 1373–1382 (2002).
30. Bement, W. M., Hasson, T., Wirth, J. A., Cheney, R. E. & Mooseker, M. S. Identification and overlapping expression of multiple unconventional myosin genes in vertebrate cell types. *Proc. Natl Acad. Sci. USA* **91**, 6549–6553 (1994).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements Thanks to G. Von Dassow and J. Canman for critical reading of this manuscript. This work was supported by grants from the National Institutes of Health to W.M.B. and R.E.C.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.M.B. (wmbement@wisc.edu).

Pilus chaperones represent a new type of protein-folding catalyst

Michael Vetsch*, Chasper Puorger*, Thomas Spirig, Ulla Grauschopf, Eilika U. Weber-Ban & Rudi Glockshuber

Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule Hönggerberg, CH-8093 Zürich, Switzerland

*These authors contributed equally to this work

Adhesive type 1 pili from uropathogenic *Escherichia coli* strains have a crucial role during infection by mediating the attachment to and potentially the invasion of host tissue. These filamentous, highly oligomeric protein complexes are assembled by the ‘chaperone–usher’ pathway¹, in which the individual pilus subunits fold in the bacterial periplasm and form stoichiometric complexes with a periplasmic chaperone molecule that is essential for pilus assembly^{2–4}. The chaperone subsequently delivers the subunits to an assembly platform (usher) in the outer membrane, which mediates subunit assembly and translocation to the cell surface^{5–8}. Here we show that the periplasmic type 1 pilus chaperone FimC binds non-native pilus subunits and accelerates folding of the subunit FimG by 100-fold. Moreover, we find that the FimC–FimG complex is formed quantitatively and very rapidly when folding of FimG is initiated in the presence of both FimC and the assembly-competent subunit FimF, even though the FimC–FimG complex is thermodynamically less stable than the FimF–FimG complex. FimC thus represents a previously unknown type of protein-folding catalyst, and simultaneously acts as a kinetic trap preventing spontaneous subunit assembly in the periplasm.

Type 1 pili from *E. coli* are composed of the main structural pilus subunit FimA, which forms the pilus rod, and subunits FimF, FimG and the adhesin FimH, which form the distal tip fibrillum⁹ (Fig. 1). X-ray structures of chaperone–subunit and subunit–subunit complexes from different adhesive pilus systems have shown that pilus subunits have an immunoglobulin-like fold lacking a β -strand^{10–13}. In chaperone–subunit complexes this incomplete fold of the pilus subunit is completed by a polypeptide segment of the chaperone, termed donor strand^{10–13}, whereas in the assembled pilus an amino-terminal extension of the adjacent pilus subunit acts as the donor strand^{12–14} (Fig. 1).

On the basis of the above complementation principles, we designed an experiment that mimics pilus subunit folding in the periplasm. To test whether a folding subunit preferentially binds to the free chaperone or directly associates with the exposed donor strand of a chaperone-bound subunit, the chemically denatured subunit FimG was refolded in the presence of both the chaperone FimC and the subunit FimF. In order to inhibit self-polymerization of FimG and FimF, we used a FimG variant (FimG_Δ) lacking the 12 N-terminal donor strand residues, and a FimF variant (FimF_Δ) elongated at its carboxy terminus by a FimF donor strand that allows intramolecular donor strand complementation (see Supplementary Fig. S1). The only complexes that can form in this

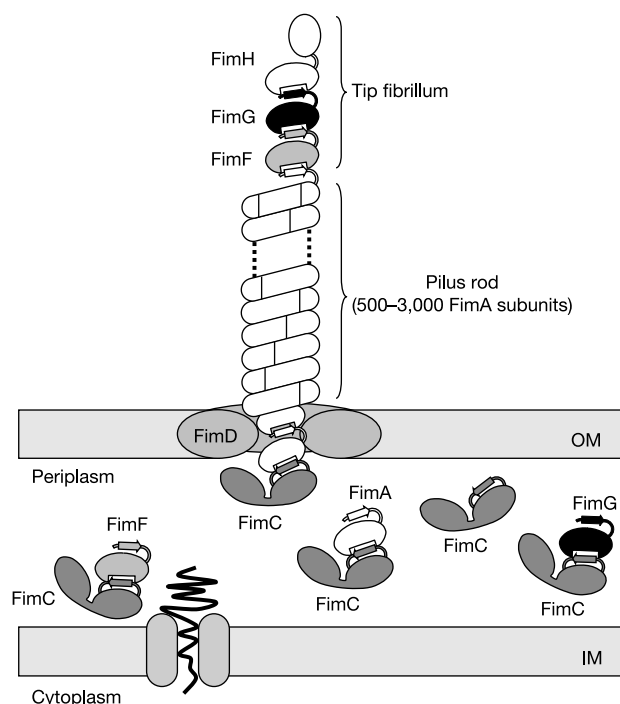


Figure 1 Schematic representation of type 1 pilus assembly according to the chaperone–usher pathway. Pilus subunits (FimA, FimF, FimG, FimH) enter the periplasm in an unfolded conformation. They subsequently form 1:1 complexes with the pilus chaperone FimC. Upon interaction with the usher FimD, which forms a pore in the outer membrane, the chaperone–subunit complexes dissociate. While the chaperone remains in the periplasm, the subunits cross the outer membrane and become incorporated into the growing pilus. IM, inner membrane; OM, outer membrane.

system are the heterodimers FimC–FimG_t and FimF_F–FimG_t (Fig. 2c). We refolded FimG_t (7.2 μ M) by dilution of the denaturant guanidinium chloride (GdmCl; see Supplementary Fig. S2) in the presence of equimolar concentrations of FimC and a 1.2-fold molar excess of FimF_F. After different reaction times, samples were withdrawn and analysed by fast anion exchange chromatography (Fig. 2a) to determine the concentrations of monomeric FimF_F, FimC–FimG_t and FimF_F–FimG_t as a function of the reaction time. Figure 2a, b shows that the FimC–FimG_t complex was formed quantitatively and very rapidly within the time required for mixing and loading of the sample onto the ion exchange column (3 min). However, FimG_t then slowly dissociated from FimC and formed a 1:1 complex with FimF_F with a half-life of 3.2 h (Fig. 2b). We find that 93% of FimG_t molecules form a complex with FimC after refolding for 3 min. After 62 h, however, a mere 0.6% of the FimG_t molecules are bound to FimC, whereas the rest are in complex with FimF_F. This shows that the FimF_F–FimG_t complex is at least -17 kJ mol^{-1} more stable than the FimC–FimG_t complex. Thus, the chaperone captures folding subunits and prevents premature pilus assembly in the periplasm while energy is preserved (at least -17 kJ mol^{-1} in the case of FimG) to promote the subsequent incorporation of subunits into the pilus. Although we observe extremely slow transfer of a chaperone-bound subunit to an assembly-competent partner subunit *in vitro*, entire pili are formed within minutes *in vivo*¹⁵, suggesting that an important role of the usher FimD is the catalysis of chaperone–subunit dissociation and subunit–subunit association.

To investigate the folding kinetics of FimG_t we used stopped-flow tryptophan fluorescence. FimG_t was refolded by 1:11 dilution of GdmCl from 2.75 to 0.25 M. We analysed folding of FimG_t alone and in the presence of either the chaperone FimC or the self-complemented pilus subunit FimF_F, both of which form stable complexes with FimG_t (Fig. 2). To avoid strong fluorescence contributions from FimC, we used the tryptophan-free FimC variant FimC_{YY}, which is fully functional and can complement FimC deficiency *in vivo* (Supplementary Fig. S5). We found that spontaneous refolding of FimG_t is a slow process that can be fit with

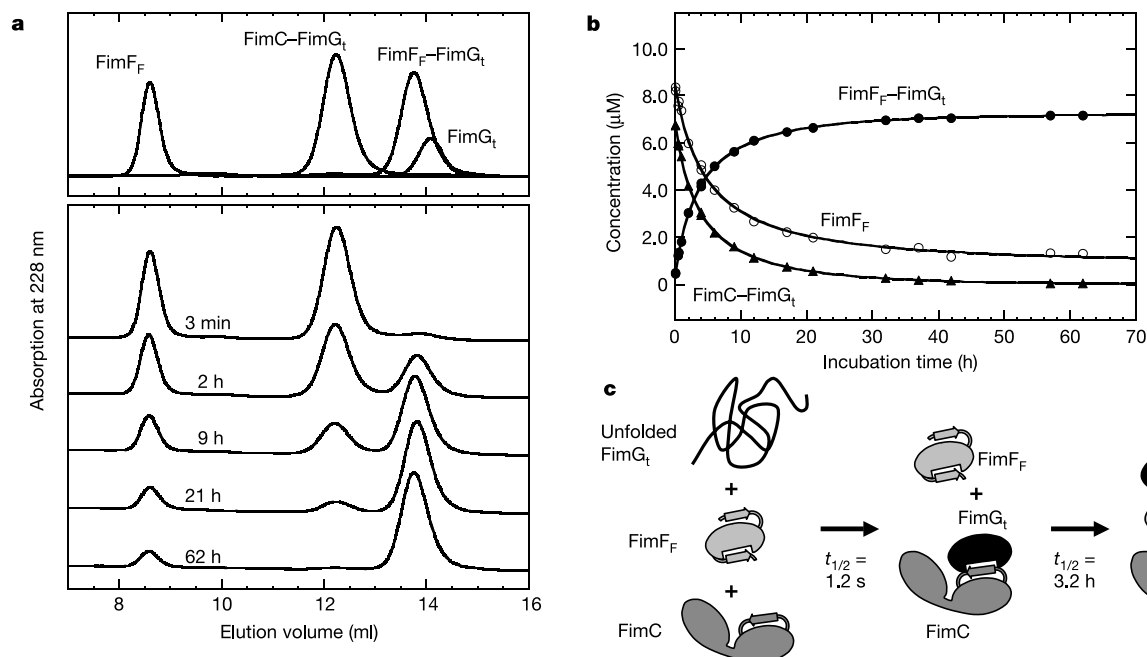


Figure 2 FimC forms transient complexes with folding pilus subunits and acts as a kinetic assembly trap. **a**, **b**, FimG_t (7.2 μ M) was refolded in presence of the chaperone FimC (7.2 μ M) and the self-complemented subunit FimF_F (8.6 μ M). After different reaction times, products were separated by anion exchange chromatography. Peak areas (**a**) were

converted to concentrations, plotted as a function of incubation time (**b**) and the resulting curves were fitted as second-order reactions (solid lines). **c**, Reaction scheme. The half-life of formation of FimC–FimG_t (1.2 s) is deduced from the experiment described in Fig. 3a.

Table 1 Relaxation rate constants and amplitudes of the formation of native FimG_t or native FimG_t*

Folding protein	Protein or peptide in the refolding buffer	$1/\tau_1$ (s ⁻¹)	A ₁ (%)	$1/\tau_2$ (s ⁻¹)	A ₂ (%)
FimG _t	FimC _{YY}	0.59 ± 0.08	55 ± 5	0.033 ± 0.007	45 ± 5
FimG _t *	FimC _{YY}	2.1 ± 0.2	78 ± 3	0.044 ± 0.011	22 ± 3
FimG _t	DS _{FimC}	0.0041 ± 0.0001	100	—	—
FimG _t *	DS _{FimC}	0.012 ± 0.004	100	—	—

A indicates amplitude. Values are mean $\pm$ s.e.m.

two exponentials ($1/\tau_1 = 0.14 \text{ s}^{-1}$, $1/\tau_2 = 0.031 \text{ s}^{-1}$; Supplementary Fig. S3a) and that is essentially not affected by FimF_F (Supplementary Fig. S3c). If, however, FimG_t is refolded in the presence of FimC_{YY}, a fast fluorescence increase is observed that is absent in the trace of spontaneous FimG_t folding. Overall, four exponentials ($1/\tau_1 = 38 \text{ s}^{-1}$, $1/\tau_2 = 1.9 \text{ s}^{-1}$, $1/\tau_3 = 0.24 \text{ s}^{-1}$, $1/\tau_4 = 0.026 \text{ s}^{-1}$) are required to fit the fluorescence trace (Supplementary Fig. S3b and Supplementary Table S1). The fast fluorescence increase of FimG_t observed only in the presence of the chaperone (see Supplementary Information Fig. S3) proves binding of the chaperone to non-native substrate (either an early folding intermediate or the completely unfolded state). The fact that two of the kinetic phases of FimG_t folding in the presence of chaperone are significantly faster than for spontaneous refolding suggests that the chaperone accelerates formation of native FimG_t. As the β -fold of FimG is complemented by the FimC donor strand in the FimC–FimG complex (Fig. 1), we next tested whether a synthetic peptide corresponding to the FimC donor strand (DS_{FimC}) accelerates folding of FimG_t. We refolded FimG_t in the presence of a 40-fold excess of DS_{FimC} and found that the peptide did not significantly affect the refolding reaction (Supplementary Fig. S3d). The same was observed for peptides corresponding to the donor strands of FimF and FimG (DS_{FimF} and DS_{FimG}, respectively; Supplementary Fig. S3e, f).

To substantiate that formation of native FimG_t, rather than formation of a folding intermediate, is accelerated by FimC, we performed interrupted refolding experiments¹⁶, which exclusively detect the formation of native FimG_t during the refolding reaction. These experiments make use of the fact that, in a mixture of native molecules and folding intermediates, native molecules unfold with a characteristic rate constant at a given denaturant concentration, whereas any folding intermediate unfolds significantly faster¹⁶. The amplitude of the slowest unfolding phase is thus directly proportional to the concentration of native molecules after the corresponding time of refolding. FimG_t was allowed to refold in the

presence of either a fivefold excess of FimC_{YY} or a 40-fold excess of DS_{FimC}. After different times of refolding, the reactions were transferred to unfolding conditions and the amplitudes of the unfolding reactions were measured. In the presence of DS_{FimC}, formation of native FimG_t can be described by a single first-order relaxation rate constant of 0.0041 s^{-1} (Fig. 3a), which is one order of magnitude slower than the slowest rate detectable in the fluorescence trace (Supplementary Table S1). Therefore, a spectroscopically silent, rate-limiting reaction leads to formation of the native FimG_t–DS_{FimC} complex. The same result was obtained for refolding in the presence of DS_{FimF} and DS_{FimG}, which, like DS_{FimC}, form specific complexes with FimG_t (Supplementary Fig. S4).

The kinetics of native FimG_t formation in the presence of FimC_{YY} are markedly different: most of the molecules ($55 \pm 5\%$; mean $\pm$ s.e.m.) reach the native state with a rate constant of $0.59 \pm 0.08 \text{ s}^{-1}$, which is more than 100-fold faster than in presence of the DS_{FimC} peptide. The remaining fraction ($45 \pm 5\%$) folds 20-fold slower ($0.033 \pm 0.007 \text{ s}^{-1}$), but still eightfold faster than in the presence of DS_{FimC} (Fig. 3a and Table 1). The rate of the slower folding reaction coincides with typical values of prolyl *cis*–*trans* isomerization¹⁷. It is thus conceivable that prolyl *cis*–*trans* isomerization is limiting the rate of folding in the fraction of GdmCl-denatured FimG_t molecules that have prolyl bonds in a non-native conformation. To test this hypothesis, we constructed FimG_t variants in which three proline residues were replaced by glycines. Expression tests revealed that the proline at position 55 is critical for the stability of FimG_t and cannot be replaced without loss of folding competence. In contrast, a variant of FimG_t with two proline residues at positions 15 and 116 replaced by glycine residues (see Supplementary Fig. S1) was expressed at the same level as FimG_t. Folding of this variant (referred to hereafter as FimG_t*) was again quantified by interrupted refolding experiments in the presence of FimC_{YY} and DS_{FimC}. In both cases, folding of FimG_t* is about threefold faster than folding of FimG_t (Fig. 3b). Notably, the fraction of slowly folding FimG_t* molecules is only 22% compared with 45% in the case of FimG_t (Fig. 3 and Table 1). This shows that non-native *cis*-prolyl bonds indeed cause the slow phase in chaperone-assisted FimG_t folding. Nevertheless, the periplasmic *E. coli* prolyl isomerase cyclophilin did not catalyse refolding of FimG_t (Supplementary Fig. S6), indicating that the prolyl bonds are not accessible to the enzyme during FimG_t folding.

The marked acceleration of folding to native FimG_t by more than 100-fold from 0.0041 s^{-1} to 0.59 s^{-1} has important implications for the kinetics of pilus assembly *in vivo*. Previous studies on the rate of protein translocation by the secretory machinery into the *E. coli* periplasm¹⁸ and the rate of pilus assembly¹⁵, which in the case of type 1 pili consists of 500–3,000 pilus subunits⁹, indicate that both translocation of an individual subunit into the periplasm and its incorporation into the growing pilus occur over the timescale of a few seconds. In contrast, uncatalysed folding of pilus subunits takes minutes (Fig. 3; see also ref. 19). The large acceleration of subunit folding by FimC is thus a prerequisite for efficient pilus biogenesis as it overcomes the kinetic bottleneck of the assembly process.

It is remarkable that FimC accelerates folding of substrate molecules in view of the fact that classical molecular chaperones enhance the efficiency but not the rate by which their substrates fold²⁰. The only known exception is the up to fourfold faster folding

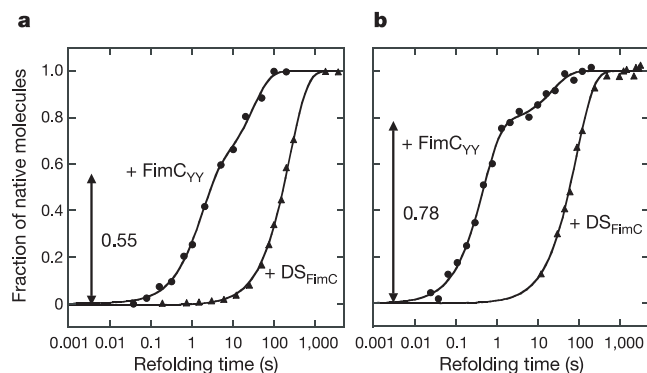


Figure 3 FimC accelerates folding of FimG_t by more than 100-fold. **a**, **b**, Formation of native FimG_t (**a**) and FimG_t* (**b**). Formation of native FimG_t and FimG_t* in the presence of FimC_{YY} (circles) is described with the sum of two exponential functions (solid lines), whereas in the presence of the DS_{FimC} peptide (triangles) a single exponential function is sufficient to fit the data (solid lines). Double-headed arrows indicate the fraction of fast-folding species in the presence of FimC_{YY}. The relaxation rate constants and the relative amplitudes are given in Table 1.

of a small set of substrates in the GroEL/GroES cavity²¹. FimC is nevertheless a typical chaperone in that it forms complexes with non-assembled substrates and also inhibits nonspecific aggregation of subunits^{19,22}. The acceleration of pilus subunit folding by FimC described here is not caused by the FimC donor strand alone but requires additional specific contacts between the folding polypeptide and the chaperone. This is reminiscent of proteins that are synthesized as precursors with pro regions that are transiently necessary for folding²³. A well-characterized example is α -lytic protease, in which the pro region markedly accelerates folding to the native protease²⁴. Whereas pro regions act as intramolecular folding catalysts, periplasmic pilus chaperones function as intermolecular catalysts that are released upon delivery of native subunits to the growing pilus for the next round of catalysis. The common feature of pro regions and periplasmic chaperones as folding catalysts may be that they both transiently become part of the tertiary structure of their substrate. So far, natural, intermolecular protein-folding catalysis has only been known for peptide bond *cis-trans* isomerization and disulphide bond formation, two prominent, rate-limiting steps in protein folding^{17,25,26}. FimC thus represents a new type of protein-folding catalyst that either accelerates certain rate-limiting steps along the folding pathway of pilus subunits or even changes the folding mechanism of its substrates entirely. □

Methods

Plasmids

For cytoplasmic expression of C-terminally (His)₆-tagged FimC the genetic sequence of mature FimC was amplified by polymerase chain reaction (PCR) from pfimC²⁷ and introduced into pET-11a (Novagen). The resulting plasmid pfimC_{His}-cyt was used to create the (His)₆-tagged variant FimC_{YY} (QuikChange, Stratagene). In order to express FimG_t* the codons for the proline residues at positions 15 and 116 of mature FimG were replaced with codons for glycine residues (QuikChange, Stratagene). For periplasmic coexpression of FimG_t* with (His)₆-tagged FimC, the gene was introduced into pfimG_t-fimC_{His}²⁸ resulting in pfimG_{15GP116Gr}-fimC_{His}. For cytoplasmic expression of FimF_t the gene encoding mature FimF was amplified by PCR from genomic DNA of *E. coli* W3110 using a 3' primer that added a short linker (DNKQ) to the FimF C terminus followed by the FimF donor strand (ADSTITIRGYVRDN). The PCR product was inserted into pET-11a yielding pfimF_t-cyt.

Proteins and peptides

FimF_t and (His)₆-tagged FimC and FimC_{YY} were expressed as soluble proteins in the cytoplasm of *E. coli* Tuner (DE3) (Novagen) carrying the plasmids pfimF_t-cyt, pfimC_{His}-cyt and pfimC_{W36YVW84YHis}-cyt, respectively. (His)₆-tagged FimC and FimC_{YY} were purified by immobilized metal affinity and cation exchange chromatography. In order to obtain FimG_t, the FimC–FimG_t complex was expressed and purified as described²⁸. (His)₆-tagged FimC was then removed by immobilized metal affinity chromatography after dissociation of the complex under denaturing conditions (2.75 M GdmCl, pH 8.0). The same protocol was applied for production of FimG_t*. FimF_t was purified by anion exchange and hydrophobic chromatography. The single disulphide bond in FimF_t was formed by air oxidation during purification, as verified by Ellman's assay after unfolding of FimF_t in 6.0 M GdmCl. The peptide DS_{FimC} was obtained from Jerini (>95% pure). Protein concentrations were determined by the extinction coefficients at 280 nm (FimC: 22,900 M⁻¹ cm⁻¹; FimC_{YY}: 14,300 M⁻¹ cm⁻¹; FimF_t: 10,900 M⁻¹ cm⁻¹; FimG_t and FimG_t*: 12,200 M⁻¹ cm⁻¹). The concentration of the peptide DS_{FimC} was measured by the extinction coefficient at 205 nm (47,400 M⁻¹ cm⁻¹).

Kinetics of FimC binding and subunit assembly

Unfolded FimG_t in 20 mM Tris/HCl, 2.75 M GdmCl, pH 8.0 was diluted 11-fold with 20 mM Tris/HCl, pH 8.0 containing FimC and FimF_t (25 °C). The final concentration was 7.2 μ M for FimG_t and FimC, and 8.6 μ M for FimF_t. At different times after mixing, samples (1 ml) were removed, cooled to 4 °C and rapidly desalted on a HiTrap Desalting column (Amersham Biotech). The protein was then immediately subjected to analytical ion exchange chromatography on a Resource Q 1 ml column (Amersham Biotech) in 10 mM Tris/HCl, pH 9.0. Bound protein was eluted with a NaCl gradient and detected at 228 nm. Using the peak areas of the individual components (FimF_t, FimC–FimG_t, FimF_t–FimG_t), of which defined amounts had been subjected to the same procedure as a reference, the amounts of these proteins and protein complexes in the reaction mixture at the various time points were determined.

Interrupted refolding experiments

All measurements were performed at 25 °C. Unfolded FimG_t or unfolded FimG_t* (in 2.75 M GdmCl) was diluted 11-fold with 20 mM Tris/HCl, pH 8.0 containing FimC_{YY} or DS_{FimC}. Final concentrations were 1 μ M for FimG_t and FimG_t*, 5 μ M for FimC_{YY} and 40 μ M for DS_{FimC}. For interrupted refolding of FimG_t in the presence of FimC_{YY}, an SX-18MV stopped-flow instrument (Applied Photophysics) was used. After various times of refolding (*t*_i), the GdmCl concentration was rapidly increased from 0.25 to 2.7 M, and the unfolding kinetics were monitored at 330 nm (excitation at 295 nm). The amplitudes corresponding to

unfolding of native FimC–FimG_t were normalized against the unfolding amplitudes of FimG_t that had been completely refolded in the presence of FimC_{YY}. This yielded the fraction of native FimG_t at *t*_i. Refolding of FimG_t* in the presence of FimC_{YY} was measured accordingly. Refolding of 1 μ M FimG_t in the presence of 40 μ M DS_{FimC} was measured in the same manner as well, except that unfolding of the native FimG_t–DS_{FimC} complex was induced by 4.1 M GdmCl, 50 mM glycine/HCl, pH 1.8, and that additional data points at 1,800 and 3,600 s were obtained by a combination of manual mixing (initiation of refolding) and stopped-flow mixing (detection of unfolding). In order to measure refolding of FimG_t* in the presence of 40 μ M DS_{FimC}, refolding was induced by manual mixing. A PTI Quantamaster QM-7 spectrofluorometer was used to record the unfolding reaction, which was induced by transferring the protein to 3.5 M GdmCl, 50 mM glycine/HCl, pH 1.8.

Received 23 June; accepted 26 July 2004; doi:10.1038/nature02891.

1. Sauer, F. G. *et al.* Chaperone-assisted pilus assembly and bacterial attachment. *Curr. Opin. Struct. Biol.* **10**, 548–556 (2000).
2. Lindberg, F., Tennent, J. M., Hultgren, S. J., Lund, B. & Normark, S. PapD, a periplasmic transport protein in P-pilus biogenesis. *J. Bacteriol.* **171**, 6052–6058 (1989).
3. Kuehn, M. J., Normark, S. & Hultgren, S. J. Immunoglobulin-like PapD chaperone caps and uncaps interactive surfaces of nascently translocated pilus subunits. *Proc. Natl Acad. Sci. USA* **88**, 10586–10590 (1991).
4. Jones, C. H. *et al.* FimC is a periplasmic PapD-like chaperone that directs assembly of type 1 pili in bacteria. *Proc. Natl Acad. Sci. USA* **90**, 8397–8401 (1993).
5. Klemm, P. & Christiansen, G. The fimD gene required for cell surface localization of *Escherichia coli* type 1 fimbriae. *Mol. Gen. Genet.* **220**, 334–338 (1990).
6. Dodson, K. W., Jacob-Dubuisson, F., Striker, R. T. & Hultgren, S. J. Outer-membrane PapC molecular usher discriminately recognizes periplasmic chaperone-pilus subunit complexes. *Proc. Natl Acad. Sci. USA* **90**, 3670–3674 (1993).
7. Thanassi, D. G. *et al.* The PapC usher forms an oligomeric channel: implications for pilus biogenesis across the outer membrane. *Proc. Natl Acad. Sci. USA* **95**, 3146–3151 (1998).
8. Saulino, E. T., Bullitt, E. & Hultgren, S. J. Snapshots of usher-mediated protein secretion and ordered pilus assembly. *Proc. Natl Acad. Sci. USA* **97**, 9240–9245 (2000).
9. Hahn, E. *et al.* Exploring the 3D molecular architecture of *Escherichia coli* type 1 pili. *J. Mol. Biol.* **323**, 845–857 (2002).
10. Choudhury, D. *et al.* X-ray structure of the FimC–FimH chaperone-adhesin complex from uropathogenic *Escherichia coli*. *Science* **285**, 1061–1066 (1999).
11. Sauer, F. G. *et al.* Structural basis of chaperone function and pilus biogenesis. *Science* **285**, 1058–1061 (1999).
12. Sauer, F. G., Pinkner, J. S., Waksman, G. & Hultgren, S. J. Chaperone priming of pilus subunits facilitates a topological transition that drives fiber formation. *Cell* **111**, 543–551 (2002).
13. Zavialov, A. V. *et al.* Structure and biogenesis of the capsular F1 antigen from *Yersinia pestis*: preserved folding energy drives fiber formation. *Cell* **113**, 587–596 (2003).
14. Zavialov, A. V. *et al.* Donor strand complementation mechanism in the biogenesis of non-pilus systems. *Mol. Microbiol.* **45**, 983–995 (2002).
15. Jacob-Dubuisson, F., Striker, R. T. & Hultgren, S. J. Chaperone-assisted self-assembly of pili independent of cellular energy. *J. Biol. Chem.* **269**, 12447–12455 (1994).
16. Schmid, F. X. Mechanism of folding of ribonuclease A. Slow refolding is a sequential reaction via structural intermediates. *Biochemistry* **22**, 4690–4696 (1983).
17. Balbach, J. & Schmid, F. X. in *Protein Folding: Frontiers in Molecular Biology* (ed. Pain, R.) (Oxford Univ. Press, Oxford, 2000).
18. Pugsley, A. P. The complete general secretory pathway in gram-negative bacteria. *Microbiol. Rev.* **57**, 50–108 (1993).
19. Vetsch, M., Seibel, P. & Glockshuber, R. Chaperone-independent folding of type 1 pilus domains. *J. Mol. Biol.* **322**, 827–840 (2002).
20. Hartl, F. U. & Hayer-Hartl, M. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**, 1852–1858 (2002).
21. Brinker, A. *et al.* Dual function of protein confinement in chaperonin-assisted protein folding. *Cell* **107**, 223–233 (2001).
22. Jones, C. H., Danese, P. N., Pinkner, J. S., Silhavy, T. J. & Hultgren, S. J. The chaperone-assisted membrane release and folding pathway is sensed by two signal transduction systems. *EMBO J.* **16**, 6394–6406 (1997).
23. Eder, J. & Fersht, A. R. Pro-sequence-assisted protein folding. *Mol. Microbiol.* **16**, 609–614 (1995).
24. Baker, D., Sohl, J. L. & Agard, D. A. A protein-folding reaction under kinetic control. *Nature* **356**, 263–265 (1992).
25. Frand, A. R., Cuozzo, J. W. & Kaiser, C. A. Pathways for protein disulphide bond formation. *Trends Cell Biol.* **10**, 203–210 (2000).
26. Schiene-Fischer, C., Habazettl, J., Schmid, F. X. & Fischer, G. The hsp70 chaperone DnaK is a secondary amide peptide bond *cis-trans* isomerase. *Nature Struct. Biol.* **9**, 419–424 (2002).
27. Hermanns, U., Seibel, P., Eggli, V. & Glockshuber, R. Characterization of FimC, a periplasmic assembly factor for biogenesis of type 1 pili in *Escherichia coli*. *Biochemistry* **39**, 11564–11570 (2000).
28. Nishiyama, M., Vetsch, M., Puorger, C., Jelezarov, I. & Glockshuber, R. Identification and characterization of the chaperone-subunit complex-binding domain from the type 1 pilus assembly platform FimD. *J. Mol. Biol.* **330**, 513–525 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Bachmann and T. Kiefhaber for their assistance and discussions; A. Fritz for technical assistance; and R. Brunisholz for Edman sequencing and MALDI–TOF mass spectrometry. This work was supported by the Schweizerische Nationalfonds and the Swiss NCCR Program in Structural Biology.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.G. (rudi@mol.biol.ethz.ch).

Direct charging of tRNA_{CUA} with pyrrolysine *in vitro* and *in vivo*

Sherry K. Blight¹*, Ross C. Larue¹*, Anirban Mahapatra¹*, David G. Longstaff¹, Edward Chang¹, Gang Zhao²†, Patrick T. Kang⁴, Kari B. Green-Church⁵, Michael K. Chan^{2,3,4} & Joseph A. Krzycki^{1,4}

¹Department of Microbiology, 484 West 12th Avenue, ²Department of Chemistry, 100 West 18th Avenue, ³Department of Biochemistry, 484 West 12th Avenue, The Ohio State University, Columbus, Ohio 43210, USA

⁴Ohio State University Biochemistry Program, 484 West 12th Avenue, The Ohio State University, Columbus, Ohio 43210, USA

⁵CCIC/Mass Spectrometry and Proteomics Facility, The Ohio State University, 116 W 19th Ave, Columbus, Ohio 43210, USA

* These authors contributed equally to this work.

† Present address: Nitto Denko Technical Corporation, 401 Jones Road, Oceanside, California 92054, USA.

Pyrrolysine is the 22nd amino acid^{1–3}. An unresolved question has been how this atypical genetically encoded residue is inserted into proteins, because all previously described naturally occurring aminoacyl-tRNA synthetases are specific for one of the 20 universally distributed amino acids. Here we establish that synthetic L-pyrrolysine is attached as a free molecule to tRNA_{CUA} by PylS, an archaeal class II aminoacyl-tRNA synthetase. PylS activates pyrrolysine with ATP and ligates pyrrolysine to tRNA_{CUA} *in vitro* in reactions specific for pyrrolysine. The addition of pyrrolysine to *Escherichia coli* cells expressing *pylT* (encoding tRNA_{CUA}) and *pylS* results in the translation of UAG *in vivo* as a sense codon. This is the first example from nature of direct aminoacylation of a tRNA with a non-canonical amino acid and shows that the genetic code of *E. coli* can be expanded to include UAG-directed pyrrolysine incorporation into proteins.

Most organisms employ UAG as a stop codon, but translation is not terminated at in-frame UAGs in some methyltransferases of methanogenic Archaea^{4,5}. Rather, these codons serve as sense codons⁶ and, as determined by crystal structure analyses, UAG encodes pyrrolysine, that is, lysine with the epsilon nitrogen in amide linkage to (4*R*, 5*R*)-4-substituted-pyrroline-5-carboxylate¹, the 22nd amino acid found to be genetically encoded in nature. A key question is whether the UAG-translating tRNA_{CUA} (ref. 2) is first charged with lysine^{2,7} and then modified to pyrrolysine for incorporation into the growing polypeptide or whether pyrrolysine is attached as the fully synthesized amino acid to tRNA_{CUA} (refs 2, 8). Here we show that the latter possibility is feasible by demonstrating the direct pyrrolysylation of tRNA_{CUA} *in vitro*. This is the first example found in nature of specific aminoacylation of a tRNA with a non-canonical amino acid. The results reported show further that the expression of only two genes, *pylT* and *pylS*, which encode tRNA_{CUA} and pyrrolysyl-tRNA synthetase, can expand the genetic code of *E. coli* to include pyrrolysine. This procedure could potentially be used to immediately expand the genetic code of any species that can incorporate exogenously added pyrrolysine.

The 4-substituent of pyrrolysine could not be initially assigned, but recent mass spectrometry with MtmB peptide fragments has provided accurate mass measurements indicating that the substituent is a methyl group (K.B.G.-C., J. A. Soares, L. Zhang, R. L. Pitsch, N. M. Kleinholz, R. Benjamin Jones, J. J. Wolff, J. Amster and J.K., unpublished observations). Crystallographic studies have also indicated that the most likely substituent at the 4-position of the pyrrolysine ring is a methyl group, and this form of L-pyrrolysine has been synthesized⁹. Recombinant PylS-His₆ was purified by Ni affinity chromatography from lysates of *Escherichia coli* expressing the recombinant *Methanosarcina barkeri* *pylS* gene modified so as to add a carboxy-terminal hexahistidine tag to the

gene product (see Supplementary Fig. S1). The tRNA pool extracted from *Methanosarcina acetivorans* or tRNA_{CUA} transcribed *in vitro* was used in charging experiments. Charged and uncharged tRNA species were separated by electrophoresis in a denaturing acid-urea polyacrylamide gel^{10,11} and tRNA_{CUA} was specifically detected by northern blotting with an oligonucleotide probe. The oligonucleotide complementary to tRNA_{CUA} could hybridize to a tRNA in the pool of tRNAs isolated from wild-type *M. acetivorans* but not to the tRNA pool from a *pylT* deletion mutant of *M. acetivorans* (A.M., A. Patel, J. Soares, R.L. and J.A.K., unpublished observations).

Both tRNA_{CUA} and aminoacyl-tRNA_{CUA} were detectable in the isolated cellular tRNA pool (Fig. 1). Alkaline hydrolysis deacylated the cellular charged species, but subsequent incubation with pyrrolysine, ATP and PylS-His₆ resulted in maximal conversion of 50% of deacylated tRNA_{CUA} to a species that migrated with the same electrophoretic mobility as the aminoacyl-tRNA_{CUA} present in the extracted cellular tRNA pool. The aminoacyl-tRNA_{CUA} synthesized *in vitro* was also sensitive to mild alkaline hydrolysis (Supplementary Fig. S2). This charged species was not formed by PylS-His₆ in the presence of a mixture of the 20 canonical amino acids each at 50 μM, or only 50 μM lysine, but was formed after the further addition of synthetic pyrrolysine (Fig. 1). PylS-His₆ conversion of tRNA_{CUA} to the charged species was therefore dependent on pyrrolysine, even in the presence of other amino acids. To determine whether pyrrolysine itself was present in the cytoplasm, we prepared a cell extract of *M. acetivorans* and separated the low-molecular-mass metabolite pool from macromolecules by ultrafiltration. The small molecule pool contained a PylS-His₆ substrate for aminoacylation of tRNA_{CUA} (Fig. 1). We also demonstrated that PylS-His₆ does not aminoacylate tRNA^{lys} in the *M. acetivorans* tRNA pool with either pyrrolysine or lysine (Supplementary Fig. S3). tRNA_{CUA} transcribed *in vitro* was also aminoacylated with synthetic pyrrolysine by PylS-His₆ (Supplementary Fig. S4) in an ATP-dependent reaction. We observed that PylS-His₆ aminoacylated with pyrrolysine a maximum of 43% of *in-vitro*-transcribed tRNA_{CUA} during the course of our experiments.

As a prerequisite of tRNA aminoacylation, an aminoacyl-adenylate and pyrophosphate are formed from the amino acid and ATP by an aminoacyl-tRNA synthetase¹². This reversible activation reaction can be assayed by the isotopic exchange of ³²P-pyrophosphate into ATP dependent on the addition of the cognate amino acid for the aminoacyl-tRNA synthetase in question¹³. PylS-His₆ catalyses a pyrophosphate-ATP isotopic exchange reaction on the addition of synthetic pyrrolysine (Fig. 2). This reaction is not dependent on the addition of cellular tRNA. Exchange activity independent of tRNA is typical of a class II aminoacyl-tRNA synthetase¹⁴. The apparent *K_m* values for pyrrolysine and ATP were 53 μM and 2 μM, respectively. The apparent *V_{max}* was 120 nmol min^{−1} per mg PylS, giving a *k_{cat}* of 6 min^{−1} for the

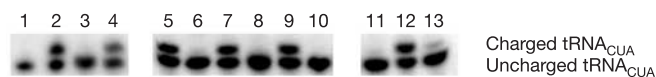


Figure 1 Aminoacylation of tRNA_{CUA} in cellular tRNA pools monitored by acid-urea gel electrophoresis and northern blotting to detect tRNA_{CUA}. Lane 1 contains the alkali-deacylated cellular tRNA pool with only uncharged tRNA_{CUA}, lane 2 contains cellular tRNA as isolated and shows charged (upper band) and uncharged (lower band) tRNA_{CUA}. Aminoacylation of tRNA_{CUA} in the deacylated cellular tRNA pool by PylS was assayed as described in Methods except that the amino-acid substrate (50 μM) was varied. Each lane represents reactions with the following: 3, no amino acid; 4, pyrrolysine; 5, pyrrolysine; 6, lysine; 7, pyrrolysine plus lysine; 8, a mixture of the 20 canonical amino acids; 9, pyrrolysine plus a mixture of the 20 canonical amino acids; 10, pyrrolysine but lacking PylS-His₆; 11, no amino acid; 12, pyrrolysine; 13, the low-molecular-mass fraction from *M. acetivorans* cell extract. Sample groups 1–4 and 5–13 are from two different experiments. Samples 5–13 are from different parts of the same gel.

exchange reaction. Incubation for as long as 30 min resulted in no detectable isotopic exchange into ATP above background in the presence of a mixture of the canonical 20 amino acids each at 100 μ M, or in the presence of 1 mM lysine.

In contrast with the inability of PylS-His₆ to synthesize lysyl-tRNA_{CUA}, we previously observed this activity with amino-terminally His-tagged PylS (His₆-PylS) as assayed by acid precipitation of tRNA ligated to radioactive lysine². However, no lysyl-tRNA synthetase activity was detectable with His₆-PylS by using the gel-shift aminoacylation assay, in agreement with a recent report⁷. In contrast, His₆-PylS does have pyrrolysyl-tRNA synthetase activity as demonstrated by the gel-shift aminoacylation assay (Supplementary Fig. S5). To determine whether PylS lacking either an N-terminal or C-terminal tag sequence acts as a lysyl- or pyrrolysyl-tRNA synthetase, we undertook the following experiments to test whether PylS allows the translation of UAG codons *in vivo* in *E. coli*, and whether this would be dependent on the presence of pyrrolysine. As a reporter of UAG translation as a sense codon, we introduced the *mtmB1* gene into *E. coli* BL21 (DE3). The *mtmB1* gene encodes the methylamine methyltransferase MtmB (refs 5, 15), in which pyrrolysine was identified¹. *E. coli* BL21 (DE3) expresses recombinant *mtmB1* with only trace amounts of the UAG read-through product, and instead primarily produces a truncated MtmB protein terminating at codon 202, the internal UAG that encodes pyrrolysine in the 452-codon *mtmB1* reading frame⁶. Plasmids were constructed bearing combinations of *mtmB1*, *pylS* and/or *pylT* under the control of the T7 promoter, and transformed into *E. coli*. Expression of these genes was induced in cells growing in the presence and the absence of exogenous 1 mM pyrrolysine. Total cellular proteins from each strain were then separated by SDS-polyacrylamide-gel electrophoresis and *mtmB1* gene products were detected by subsequent immunoblotting with polyclonal antibody specific for purified *M. barkeri* MtmB (ref. 6). All strains produced the amber-truncated product of *mtmB1* as a 23-kDa protein; however, the strain expressing *pylT*, *pylS* and *mtmB1* further expressed large amounts of 50-kDa MtmB, showing UAG read-through dependent on the presence of pyrrolysine. The pool of

amino acids in *E. coli* did not support the synthesis of an aminoacyl-tRNA_{CUA} that could efficiently translate the UAG codon in *mtmB1*. The requirement for pyrrolysine could not be replaced by 1 mM lysine or a mixture of the 20 canonical amino acids each at 1 mM. Translation of the *mtmB1* UAG codon dependent on pyrrolysine was further dependent on the expression of *pylS* and *pylT*, as demonstrated by strains transformed with expression vector containing only *pylT* (Fig. 3) or only *pylS* (Supplementary Fig. S6).

To confirm that synthetic pyrrolysine is incorporated at the UAG-encoded position of *mtmB1* by *E. coli* transformed with *pylT* and *pylS*, the insoluble recombinant full-length MtmB protein was partly purified by differential centrifugation and solubilization with urea. After SDS gel electrophoresis, full-length recombinant MtmB was subjected to in-gel digestion with chymotrypsin. An *m/z* 791.5²⁺ ion was identified, which corresponds to the predicted *m/z* 791.4²⁺ of the MtmB fragment AGRPGM_{ox}GVXGPETSL (residues 194–208), where X is the UAG-encoded residue with the predicted mass of synthetic pyrrolysine. Collision-induced dissociation mass spectrometry confirmed the sequence, and the mass of the UAG-encoded residue was ascertained as 237.2 Da (Supplementary Fig. S7, Supplementary Table S1). The predicted molecular mass of the synthetic pyrrolysyl residue is 237.16 Da, thereby confirming that expression of *pylT* and *pylS* is sufficient to expand the genetic code of *E. coli* to include exogenous pyrrolysine. The amino acid might enter the cell because its amide bond allows recognition by a broad-spectrum peptide transporter such as DppA (ref. 16). The synthesis of full-length MtmB further indicates that the *E. coli* translation factor EF-Tu binds pyrrolysyl-tRNA_{CUA} within a thermodynamic range allowing incorporation into protein during translation¹⁷.

The current data indicate that pyrrolysine is encoded in DNA using the general mechanism employed for the common set of 20 amino acids. Direct charging of pyrrolysine onto tRNA contrasts with selenocysteine, a genetically encoded non-canonical amino acid synthesized only on tRNA^{18,19}. Several systems have been recently developed to expand and manipulate the genetic code^{20–23} to generate recombinant proteins containing unnatural amino

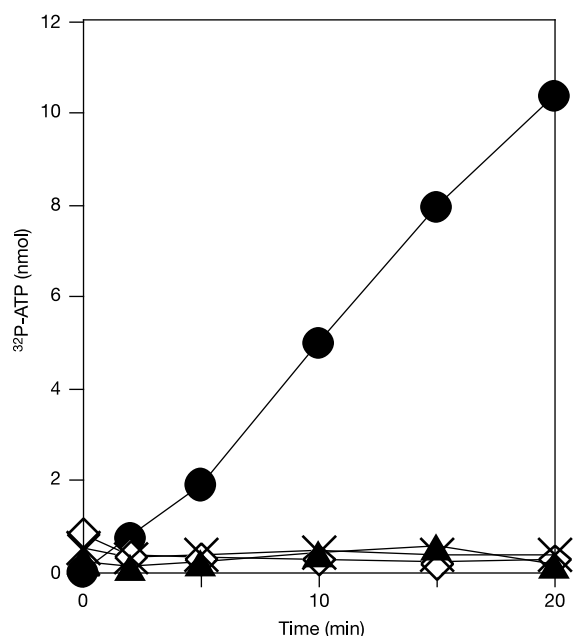


Figure 2 Dependence on pyrrolysine of the ³²P-ATP exchange reaction mediated by PylS-His₆. Aliquots were removed at the time points shown for estimation of the amount of ³²P-ATP formed. In the figure are illustrated results from averaged duplicate reactions that were either complete (filled circles) or lacked PylS-His₆ (open diamonds), ATP (filled triangles) or synthetic pyrrolysine (crosses).

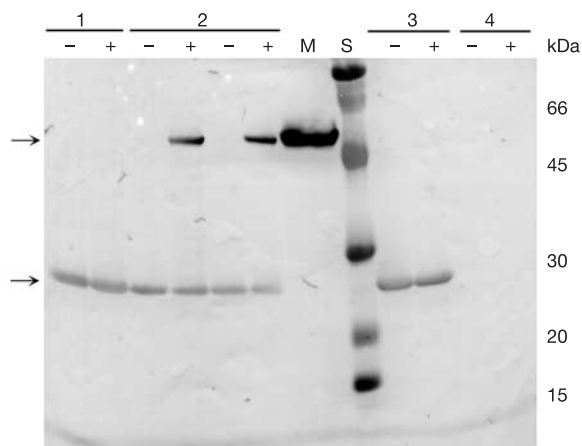


Figure 3 Anti-MtmB immunoblot of cell extracts of *E. coli* strains testing the suppression of the UAG codon in *mtmB1*. The samples loaded were from strains transformed with the following: 1, pEC02, bearing *mtmB1* and *pylT*; 2, pEC03, bearing *mtmB1*, *pylT* and *pylS*; 3, pEC01, bearing *mtmB1*; 4, pET-Duet, the vector from which the pEC plasmids are derived. Recombinant gene expression was induced in the presence (+) and absence (–) of 1 mM exogenous synthetic pyrrolysine. Also present on the immunoblot is purified 50-kDa (full-length) MtmB (M) and dye-coloured protein standards (S) for molecular mass determination; molecular masses are indicated at the right. The arrows at the left point to full-length MtmB (upper arrow) and truncated MtmB (lower arrow) made, respectively, by UAG readthrough or UAG termination during *mtmB1* expression.

acids. By adding *pylS* and *pylT* genes, it should now be possible to generate proteins with the 22nd amino acid incorporated at UAG-targeted sites in any species that can incorporate added pyrrolysine, thereby adding a unique natural amino acid with electrophilic properties. We are now focusing on the pyrrolysine biosynthetic pathway, which offers the possibility of also adding genes that will generate pyrrolysine internally in recombinant organisms. □

Methods

Recombinant proteins

The *M. barkeri* MS *pylS* gene was amplified by polymerase chain reaction (PCR) from isolated genomic DNA²⁴ and cloned into pET 22b (Novagen, Madison, Wisconsin) to create ppylSH6, which produced PylS with a hexahistidine tag at the C terminus (PylS-His₆) in *E. coli* BL21 (DE3) (Stratagene, La Jolla, California). PylS-His₆ was isolated from cell extracts in 20 mM sodium phosphate, 500 mM NaCl, 10 mM imidazole pH 7.4, using a Ni-activated trap chelating HP column (Amersham Biosciences Corp., Piscataway, New Jersey). PylS-His₆ eluted at 240 mM imidazole during the application of 10–500 mM imidazole in the same buffer to the column. His₆-PylS (ref. 2) with a hexahistidine N-terminal tag was used as a partly pure fraction from a nickel-affinity column. Control experiments indicated no pyrrolysyl-tRNA synthetase activity in untransformed *E. coli*.

The *lysS* gene was PCR amplified from *M. barkeri* MS genomic DNA for the recombinant expression of *lysS* with an N-terminal hexahistidine sequence (His₆-LysS) that eluted at about 130 mM imidazole from the nickel-affinity column.

PylS substrates

L-Pyrrolysine was synthesized and characterized with the use of ¹³C and ¹H NMR⁹. TLC analysis revealed no other amino acids. The pyrrolysine used in charging experiments was further analysed by electrospray mass spectrometry and revealed two predominant peaks with *m/z* 256.16 (M + H) and 278.14 (M + Na), where M is L-pyrrolysine. The cellular tRNA pool was isolated²⁵ from *M. acetivorans* C2A (OD₆₀₀ 0.6–0.7) growing on trimethylamine at 37 °C in DSM 304 medium²⁶ because this species is easily lysed. Agarose-gel electrophoresis indicated that 30% of the ethidium bromide staining material in the preparation was tRNA. *M. barkeri* Fusaro tRNA_{CUA} transcribed *in vitro* was produced with the DNA template described previously² and the T7-MEGAshortscript transcription kit (Ambion Inc., Austin, Texas).

The low-molecular-mass cell fraction used in aminoacylation reactions was the supernatant of French-pressed trimethylamine-grown *M. acetivorans* (27 g in 30 ml 50 mM MOPS pH 7.0) filtered with a 3-kDa Amicon Centricon apparatus (Millipore, Billerica, Massachusetts), and evaporated to dryness before resuspension in 2 ml doubly distilled water.

Aminoacylation and pyrophosphate exchange assays

The assay for aminoacylation of tRNA_{CUA} (in a volume of 25 µl) contained 0.8–1.7 µM purified PylS-His₆, 50 mM KCl, 1 mM MgCl₂, 5 mM ATP, 0.5 mM dithiothreitol and 50 µM synthetic pyrrolysine in 10 mM HEPES buffer pH 7.2, and 8 µg *M. acetivorans* tRNA pool preparation or 40 nM of tRNA_{CUA} transcript. The reaction was terminated after 5–30 min at 37 °C with an equal volume of 0.3 M sodium acetate, 8 M urea pH 5.0. Charged and uncharged tRNA were separated by acid-urea acrylamide gel electrophoresis²³, blotted to nitrocellulose and probed with a 5' ³²P-end-labelled, 72-base oligonucleotide complementary to tRNA_{CUA}. Radioactivity was analysed with a STORM Phosphorimager (Amersham Biosciences).

The pyrophosphate exchange reaction was performed after the method of ref. 13. The 100–200-µl reactions incubated at 37 °C typically contained 0.3–1 µM PylS-His₆, 10 mM MgCl₂, 25 mM KCl, 1 mM KF, 4 mM dithiothreitol, 2 mM ATP, 100 µM pyrrolysine, 2 mM ³²P-PP_i (4–10 d.p.m. pmol⁻¹; PerkinElmer, Boston, Massachusetts) in 20 mM HEPES-KOH pH 7.2.

Pyrrolysine-dependent amber suppression in *E. coli*

The *M. barkeri* MS *mtmB1* (GenBank accession number AF013713) was removed with *NdeI* and *EcoRV* from plasmid pCJ09 (ref. 6, and ligated into MCS2 of pET-Duet to create pEC01 (Novagen). The *pylS* and *pylT* genes were PCR amplified from genomic *M. acetivorans* DNA (GenBank accession number NC 003552) and *pylT* cloned into the *XbaI* site directly upstream of MCS1 in pEC01 to create pEC02. The *pylS* gene was inserted into the *NcoI* and *BamHI* sites of MCS1 of pEC02 to create pEC03. The *pylT XbaI* fragment was excised from pEC03 to create pEC05. All constructs were confirmed by restriction mapping and sequencing. To test for amber suppression, overnight cultures were grown in Luria-Bertani broth (3 ml) with 100 µg ml⁻¹ ampicillin. Subsequently, 200 µl was inoculated into 1 ml fresh medium and grown to an OD₆₀₀ of 0.6. The culture (100 µl) was then transferred to a polypropylene tube and induced for 4 h with 1 mM isopropyl β-D-thiogalactoside in the presence or absence of 1 mM pyrrolysine. The *mtmB1* gene products in equivalent amounts of lysates were then analysed by immunoblotting of a SDS 12.5% polyacrylamide gel with affinity-purified rabbit anti-MtmB antibody⁶. The Rainbow Molecular Weight markers (Amersham Biosciences) were used.

To isolate recombinant MtmB for mass spectrometry, *E. coli* bearing pEC03 (10 ml) was used with 0.75 mM pyrrolysine. The *mtmB1* gene products were inclusion bodies.

Sequentially washing the pellet from a French-pressed cell lysate with 0, 1, 3, 5 and 7 M urea in 50 mM MOPS pH 7 yielded purified *mtmB1* gene products in 7 M urea; these were separated by SDS gel electrophoresis. The 50-kDa MtmB was subjected to in-gel chymotrypsin digestion and peptide sequencing by tandem mass spectrometry²⁷.

Received 30 April; accepted 30 July 2004; doi:10.1038/nature02895.

Published online 25 August 2004.

- Hao, B. *et al.* A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* **296**, 1462–1466 (2002).
- Srinivasan, G., James, C. M. & Krzycki, J. A. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* **296**, 1459–1462 (2002).
- Atkins, J. F. & Gesteland, R. The 22nd amino acid. *Science* **296**, 1409–1411 (2002).
- Paul, L., Ferguson, D. J. & Krzycki, J. A. The trimethylamine methyltransferase gene and multiple dimethylamine methyltransferase genes of *Methanosarcina barkeri* contain in-frame and read-through amber codons. *J. Bacteriol.* **182**, 2520–2529 (2000).
- Burke, S. A., Lo, S. L. & Krzycki, J. A. Clustered genes encoding the methyltransferases of methanogenesis from monomethylamine. *J. Bacteriol.* **180**, 3432–3440 (1998).
- James, C. M., Ferguson, T. K., Leykam, J. F. & Krzycki, J. A. The amber codon in the gene encoding the monomethylamine methyltransferase isolated from *Methanosarcina barkeri* is translated as a sense codon. *J. Biol. Chem.* **276**, 34252–34258 (2001).
- Polcarpo, C. *et al.* Activation of the pyrrolysine suppressor tRNA requires formation of a ternary complex with class I and class II lysyl-tRNA synthetases. *Mol. Cell* **12**, 287–294 (2003).
- Ibba, M. & Söll, D. Aminoacyl-tRNAs: setting the limits of the genetic code. *Genes Dev.* **18**, 731–738 (2004).
- Hao, B. *et al.* Reactivity and chemical synthesis of L-pyrrolysine – the 22nd amino acid. *Chem. Biol.* (in the press).
- Ho, Y. S. & Kan, Y. W. *In vivo* aminoacylation of human and *Xenopus* suppressor tRNAs constructed by site-specific mutagenesis. *Proc. Natl Acad. Sci. USA* **84**, 2185–2188 (1987).
- Varshney, U., Lee, C. P. & RajBhandary, U. L. Direct analysis of aminoacylation levels of tRNAs *in vivo*. Application to studying recognition of *Escherichia coli* initiator tRNA mutants by glutamyl-tRNA synthetase. *J. Biol. Chem.* **266**, 24712–24718 (1991).
- Schimmel, P. & Söll, D. Aminoacyl-tRNA synthetases: general features and recognition of transfer RNAs. *Annu. Rev. Biochem.* **48**, 601–648 (1979).
- Cole, F. & Schimmel, P. R. On the rate law and mechanism of the adenosine triphosphate-pyrophosphate isotope exchange reaction of aminoacyl-transfer ribonucleic acid synthetases. *Biochemistry* **9**, 480–489 (1970).
- Francklyn, C., Perona, J. J., Puetz, J. & Hou, Y. M. Aminoacyl-tRNA synthetases: versatile players in the changing theater of translation. *RNA* **8**, 1363–1372 (2002).
- Burke, S. A. & Krzycki, J. A. Reconstitution of monomethylamine:coenzyme M methyl transfer with a corrinoid protein and two methyltransferases purified from *Methanosarcina barkeri*. *J. Biol. Chem.* **272**, 16570–16577 (1997).
- Smith, M. W., Tyreman, D. R., Payne, G. M., Marshall, N. J. & Payne, J. W. Substrate specificity of the periplasmic dipeptide-binding protein from *Escherichia coli*: experimental basis for the design of peptide prodrugs. *Microbiology* **145**, 2891–2901 (1999).
- LaRiviere, F. J., Wolfson, A. D. & Uhlenbeck, O. C. Uniform binding of aminoacyl-tRNAs to elongation factor Tu by thermodynamic compensation. *Science* **294**, 165–168 (2001).
- Stadtman, T. C. Selenocysteine. *Annu. Rev. Biochem.* **65**, 83–100 (1996).
- Commans, S. & Bock, A. Selenocysteine inserting tRNAs: an overview. *FEMS Microbiol. Rev.* **23**, 335–351 (1999).
- Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of *Escherichia coli*. *Science* **292**, 498–500 (2001).
- Döring, V. *et al.* Enlarging the amino acid set of *Escherichia coli* by infiltration of the valine coding pathway. *Science* **292**, 501–504 (2001).
- Kiick, K. L., Weberskirch, R. & Tirrell, D. A. Identification of an expanded set of translationally active methionine analogues in *Escherichia coli*. *FEBS Lett.* **502**, 25–30 (2001).
- Chin, J. W. *et al.* An expanded eukaryotic genetic code. *Science* **301**, 964–967 (2003).
- Paul, L. & Krzycki, J. A. Sequence and transcript analysis of a novel *Methanosarcina barkeri* methyltransferase II homolog and its associated corrinoid protein homologous to methionine synthase. *J. Bacteriol.* **178**, 6599–6607 (1996).
- Jester, B. C., Levengood, J. D., Roy, H., Ibba, M. & Devine, K. M. Nonorthologous replacement of lysyl-tRNA synthetase prevents addition of lysine analogues to the genetic code. *Proc. Natl Acad. Sci. USA* **100**, 14351–14356 (2003).
- Sowers, K. R. & Schreier, H. J. in *Archaea, a Laboratory Manual* (eds Sowers, K. R. & Schreier, H. J.) 459–489 (Cold Spring Harbor Laboratory Press, Plainview, New York, 1995).
- Wilm, M. *et al.* Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**, 466–469 (1996).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Reeve, C. Daniels and J. Soares for discussions, R. L. Pitsch and N. M. Kleinholz at the Ohio State University's Campus Chemical Instrumentation Center for mass spectroscopic analyses, G. Srinivasan for construction of the *lysS* expression plasmid, and S. B. Smith for mass culture of methanogens.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.A.K. (krzycki.1@osu.edu).

nature insight

RNA interference



The most familiar role for RNA is as a relatively passive intermediary in the translation of information from genes into proteins. But other functions for this versatile molecule have been emerging. This Insight explores the surprising recent discovery that RNA can actively regulate gene expression.

RNA interference or RNAi is a remarkable process whereby small noncoding RNAs silence specific genes. Although the first hints about this silencing process were seen decades ago, the real breakthrough in mechanistic understanding came in 1998. A landmark paper by Fire and colleagues (*Nature* **391**, 806–811; 1998) showed that the trigger for gene silencing is double-stranded RNA. Since then, other components of the RNAi machinery have been identified at a startling rate, although the picture is still incomplete.

RNAi is a conserved mechanism that pervades the biological world (budding yeast being a notable exception). It was first observed in plants in the guise of a mysterious immune response to viral pathogens. But RNAi is more than just a response to exogenous genetic material. Small RNAs termed microRNAs regulate gene expression in organisms ranging from nematode to man. Hundreds of microRNAs have been identified *in silico*, and we are beginning to understand their diverse functions in development and physiology.

Another biological role of RNAi is in heterochromatin silencing. By silencing transcripts generated from repeat sequences, RNAi guides heterochromatin formation and represses transposable elements and other foreign DNA integrated into the genome.

As well as expanding our appreciation of how genes are regulated, RNAi has been harnessed as an experimental tool to explore gene function, is revolutionizing our ability to perform large-scale genetic screens, and even shows therapeutic potential. We hope that these reviews capture the excitement and promise of this fast-moving field. We are grateful to the authors for their contributions and to the reviewers for their valuable input.

We are pleased to acknowledge the financial support of Merck & Co., Inc. and Alnylam Pharmaceuticals Inc. in producing this Insight. As always, *Nature* carries sole responsibility for editorial content and peer review.

Alex Eccleston and Angela K Eggleston Senior Editors

introduction

338 Revealing the world of RNA interference

C. C. Mello
& D. Conte Jr

review articles

343 Mechanisms of gene silencing by double-stranded RNA

G. Meister & T. Tuschl

350 The functions of animal microRNAs

V. Ambros

356 RNA silencing in plants

D. Baulcombe

364 The role of RNA interference in heterochromatic silencing

Z. Lippman
& R. Martienssen

371 Unlocking the potential of the human genome with RNA interference

G. J. Hannon
& J. J. Rossi

Editor, *Nature*: Philip Campbell
Insights Publisher: Sarah Greaves
Insights Editor: Lesley Anson
Editorial Assistant: Timothy Gibbs
Production Editor: Lucy Odling-Smee

Art Editor: Martin Harrison
Layouts: Alisdair Macdonald
Nadia Rooney
Diagrams: Nik Spencer

Sponsorship: Claire Hines
Production: Sue Gray
Marketing: Claire Aspinall

Revealing the world of RNA interference

Craig C. Mello^{1,2} & Darryl Conte Jr²

¹Howard Hughes Medical Institute and ²Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA (e-mail: craig.mello@umassmed.edu)

The recent discoveries of RNA interference and related RNA silencing pathways have revolutionized our understanding of gene regulation. RNA interference has been used as a research tool to control the expression of specific genes in numerous experimental organisms and has potential as a therapeutic strategy to reduce the expression of problem genes. At the heart of RNA interference lies a remarkable RNA processing mechanism that is now known to underlie many distinct biological phenomena.

The term 'RNA world' was first coined to describe a hypothetical stage in the evolution of life some four billion years ago when RNA may have been the genetic material and catalyst for emerging life on Earth^{1,2}. This original RNA world, if it ever existed on Earth, is long gone. But this Insight deals with a process that reflects an RNA world that is alive and thriving within our cells — RNA silencing or RNA interference (RNAi). When exposed to foreign genetic material (RNA or DNA), many organisms mount highly specific counter attacks to silence the invading nucleic-acid sequences before these sequences can integrate into the host genome or subvert cellular processes. At the heart of these sequence-directed immunity mechanisms is double-stranded RNA (dsRNA). Interestingly, dsRNA does more than help to defend cells against foreign nucleic acids — it also guides endogenous developmental gene regulation, and can even control the modification of cellular DNA and associated chromatin. In some organisms, RNAi signals are transmitted horizontally between cells and, in certain cases, vertically through the germ line from one generation to the next. The reviews in this Insight show our progress in understanding the mechanisms that underlie RNA-mediated gene regulation in plants and animals, and detail current efforts to harness this mechanism as a research tool and potential therapy. Here we introduce the world of RNAi, and provide a brief overview of this rapidly growing field.

Discovering the trigger

Crucial to understanding a gene-silencing mechanism such as RNAi is knowing how to trigger it. This is important from the theoretical perspective of understanding a remarkable biological response (see review in this issue by Meister and Tuschl, page 343); but it also has obvious practical ramifications for using the silencing mechanism as an experimental tool (see review in this issue by Hannon and Rossi, page 371). The observation by Fire *et al.*³ that dsRNA is a potent trigger for RNAi in the nematode *Caenorhabditis elegans* (Fig. 1) was important because it immediately suggested a simple approach for efficient induction of gene silencing in *C. elegans* and other organisms, and accelerated the discovery of a unifying mechanism that underlies a host of cellular and developmental pathways. However, there were substantial barriers to the acceptance of the idea that dsRNA could trigger sequence-specific gene silencing.

First, at the time, dsRNA was thought to be a nonspecific silencing agent that triggers a general destruction of messenger RNAs and the complete suppression of protein

translation in mammalian cells^{4,5}. Second, dsRNA is energetically stable and inherently incapable of further specific Watson–Crick base pairing. So a model in which dsRNA activates sequence-specific silencing implies the existence of cellular mechanisms for unwinding the dsRNA and promoting the search for complementary base-pairing partners among the vast pool of cellular nucleic-acid sequences. Hypotheses that require a paradigm shift and depend on the existence of a whole set of hitherto unknown activities are rarely appealing.

So why was dsRNA proposed as a trigger for RNAi and why was this idea so rapidly accepted? To answer this question we must make a brief historical digression. In 1995, Guo and Kemphues⁶ attempted to use RNA complementary to the *C. elegans par-1* mRNA to block *par-1* expression. This technique is known as 'antisense-mediated silencing', whereby large amounts of a nucleic acid whose sequence is complementary to the target messenger RNA are delivered into the cytoplasm of a cell. Base pairing between the 'sense' mRNA sequence and the complementary 'antisense' interfering nucleic acid is thought to passively block the processing or translation of mRNA, or result in the recruitment of nucleases that promote mRNA destruction^{7,8}. To their surprise, Guo and Kemphues found that both the antisense and the control sense RNA preparations induced silencing. Sense RNA is identical to the mRNA and so cannot base pair with the mRNA to cause interference, raising the question of how this RNA could induce silencing. Was an active silencing response being triggered against the foreign RNA, regardless of its polarity? Or was the silencing apparently induced by sense RNA actually mediated by antisense RNA? (Antisense RNA was known to contaminate the type of *in vitro* transcription products used in these assays.) Despite confusion about the nature of the RNA that triggered the phenomenon, this so-called antisense-mediated silencing method continued to be used to silence genes in *C. elegans*.

More surprises were in store. While using this antisense technique to silence *C. elegans* genes, we were amazed to find that the silencing effect could be transmitted in the germ line³. A remarkably potent silencing signal could be passed through the sperm or the egg for up to several generations^{3,9}. Equally remarkable, the silencing effect could also spread from tissue to tissue within the injected animal³. Taken together, the apparent lack of strand specificity, the remarkable potency of the RNA trigger, and the systemic spread and inheritance properties of the silencing phenomenon prompted the creation of a new term, RNAi¹⁰. Importantly, the properties of RNAi demanded the existence of cellular

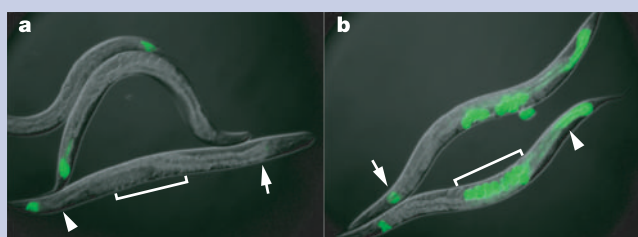


Figure 1 RNAi in *C. elegans*. Silencing of a green fluorescent protein (GFP) reporter in *C. elegans* occurs when animals feed on bacteria expressing GFP dsRNA (**a**) but not in animals that are defective for RNAi (**b**). Note that silencing occurs throughout the body of the animal, with the exception of a few cells in the tail that express some residual GFP. The signal is lost in intestinal cells near the tail (arrowhead) as well as near the head (arrow). The lack of GFP-positive embryos in **a** (bracketed region) demonstrates the systemic spread and inheritance of silencing.

mechanisms that initiate and amplify the silencing signal, and led us to suggest that the RNAi mechanism represents an active organismal response to foreign RNA³.

Although our initial models saw dsRNA as an intermediate in the amplification of the silencing signal, Fire³ suggested that dsRNA, which is often encountered by cells during viral infection, might itself be the initial trigger. In this model, instead of antisense RNA passively initiating silencing by pairing with the target mRNA, the presence of low concentrations of both sense and antisense strands in the RNA preparation was proposed to result in small amounts of dsRNA: on introduction into the animal, this dsRNA could be recognized as foreign, thereby activating cellular amplification and inheritance mechanisms. Because it was possible to produce and purify *in vitro* synthesized RNA and introduce it directly into *C. elegans* without the need for transgene-driven expression, this theory was easily tested. dsRNA proved to be an extremely potent activator of RNAi — at least 10-fold and perhaps 100-fold more effective than purified preparations of single-stranded RNA³.

Taking the biological world by storm

With the discovery of an extremely potent trigger for RNAi, it became possible to expose large populations of animals to dsRNA: animals were soaked in dsRNA¹¹ or given food containing bacterially expressed dsRNA^{12,13}. By facilitating genetic screens, these methods led to the identification of many *C. elegans* genes required for RNAi¹⁴. Comparison of the *C. elegans* genes required for RNAi to genes required for gene silencing in *Drosophila*^{15,16}, plants¹⁷ and fungi¹⁸ confirmed that the silencing phenomena known variously as post-transcriptional gene silencing (PTGS)¹⁹, co-suppression²⁰, quelling²¹ and RNAi, share a common underlying mechanism that reflects an ancient origin in a common ancestor of fungi, plants and animals. This realization was followed by a flurry of exciting results: dsRNA was shown to induce silencing in *Drosophila*²², and in a host of other organisms including organisms that were otherwise unsuited to genetic analysis^{23,24}. Small RNAs were shown to be produced in plants undergoing PTGS²⁵, and were identified as the common currency of RNA silencing pathways^{26–28} (see review in this issue by Baulcombe, page 356). The dsRNA-processing enzyme Dicer²⁹ was found to produce these small RNAs, now called short interfering RNAs (siRNAs). Synthetic RNAs engineered to look like the products of Dicer were shown to induce sequence-specific gene silencing in human cells without initiating the nonspecific gene silencing pathways³⁰. A class of natural hairpin dsRNAs^{31,32}, now called microRNAs (miRNAs; see review in this issue by Ambros, page 350), was shown to be processed by Dicer^{33–35} and to function together with RDE-1 homologues³⁵, thereby linking the RNAi machinery to a natural developmental gene regulatory mechanism. Finally, more recently, the RNAi machinery was linked to chromatin regulation in yeast³⁶, and to chromosomal rearrangement during development of the somatic macronucleus in

*Tetrahymena*³⁷. These and other breakthroughs united previously disparate fields by identifying a common core mechanism that involves the processing of dsRNA into small RNA-silencing guides (Fig. 2). In short, dsRNA had taken the biological world by storm.

Other silencing triggers

Although it was clear that dsRNA was important either as a silencing trigger or as an intermediate in all the RNAi-related silencing pathways, it was not known whether other stimuli (besides dsRNA) could trigger silencing. For example, silencing in response to a DNA transgene could still involve a dsRNA trigger: the transgene might integrate itself into the genome in such a way that a nearby promoter, or an inverted copy of the transgene itself, leads to the production of dsRNA, which could in turn enter directly into the RNAi pathway. Consistent with this idea, transgenes engineered to express both sense and antisense strands of a gene in plants can lead to efficient silencing, which is more reproducible and robust than that achieved by transgenes expressing either strand alone³⁸.

But several lines of evidence suggest that transgenes can trigger silencing through mechanisms not involving a dsRNA trigger (Fig. 2). A key gene family involved in silencing pathways in plants^{39,40}, fungi⁴¹ and *C. elegans*^{42,43} contains genes that encode putative cellular RNA-dependent RNA polymerases (RdRPs; also known as RDRs). Members of this family of proteins were identified in forward genetic screens — whereby mutant genes are isolated from an organism showing abnormal phenotypic characteristics — as factors required for co-suppression in plants and quelling in *Neurospora*. (Co-suppression results from post-transcriptional silencing of both a transgene and the endogenous copies of the corresponding cellular gene.) Interestingly, although cellular RdRP genes were required for transgene-mediated co-suppression in plants^{39,40}, they were not essential for virus-induced silencing of a transgene³⁹, presumably because the virus provides its own viral RNA polymerase. Furthermore, RdRPs have been shown to direct primer-independent synthesis of complementary RNA^{44,45}. Together, these findings suggest that the transgene or its single-stranded mRNA products could be the original stimulus for co-suppression and quelling. In this type of silencing, the RdRP somehow recognizes transgene products as abnormal or 'aberrant' and subsequently converts this initial silencing trigger into dsRNA^{46,47}. In this case, the dsRNA is an intermediate in the silencing pathway rather than the trigger. The RdRP-derived dsRNA is then likely to be processed by Dicer and to enter downstream silencing complexes that are similar, or identical, to those formed in response to a dsRNA trigger.

But how might the transgene mRNA be recognized as foreign? The answer to this question is not known. Hence, this 'aberrant transcript' model has, perhaps undeservedly, received little attention of late. One possibility discussed by Baulcombe (review in this issue, page 356) is that high levels of expression of the transgene mRNA leads to the accumulation of mRNA-processing defects (for example, non-polyadenylated transcripts) that are somehow recognized by the RdRP. Alternatively, the transgene DNA or the chromatin itself may be 'marked' for silencing by the cell. When initially delivered to cells, the transgene DNA could be recognized as foreign owing to its lack of associated proteins. During the rapid assembly of naked DNA into chromatin⁴⁸, the host cell may, in self-defence, somehow mark the transgene chromatin so that RdRP is recruited. RdRP acting on nascent transcripts could then result in dsRNA formation and subsequent silencing. Consistent with this possibility, fission yeast RdRP was found to physically associate with silent heterochromatin³⁶. Despite the mysterious nature of the silencing mark recognized by RdRPs, it seems likely that, at least in some cases, RdRPs may produce dsRNA that functions as an intermediate rather than as the primary trigger for silencing.

Genetic studies suggest that distinct silencing triggers may also exist in *C. elegans*. Both RDE-1 and the dsRNA-binding protein RDE-4 (ref. 49) are essential for mediating the silencing induced by

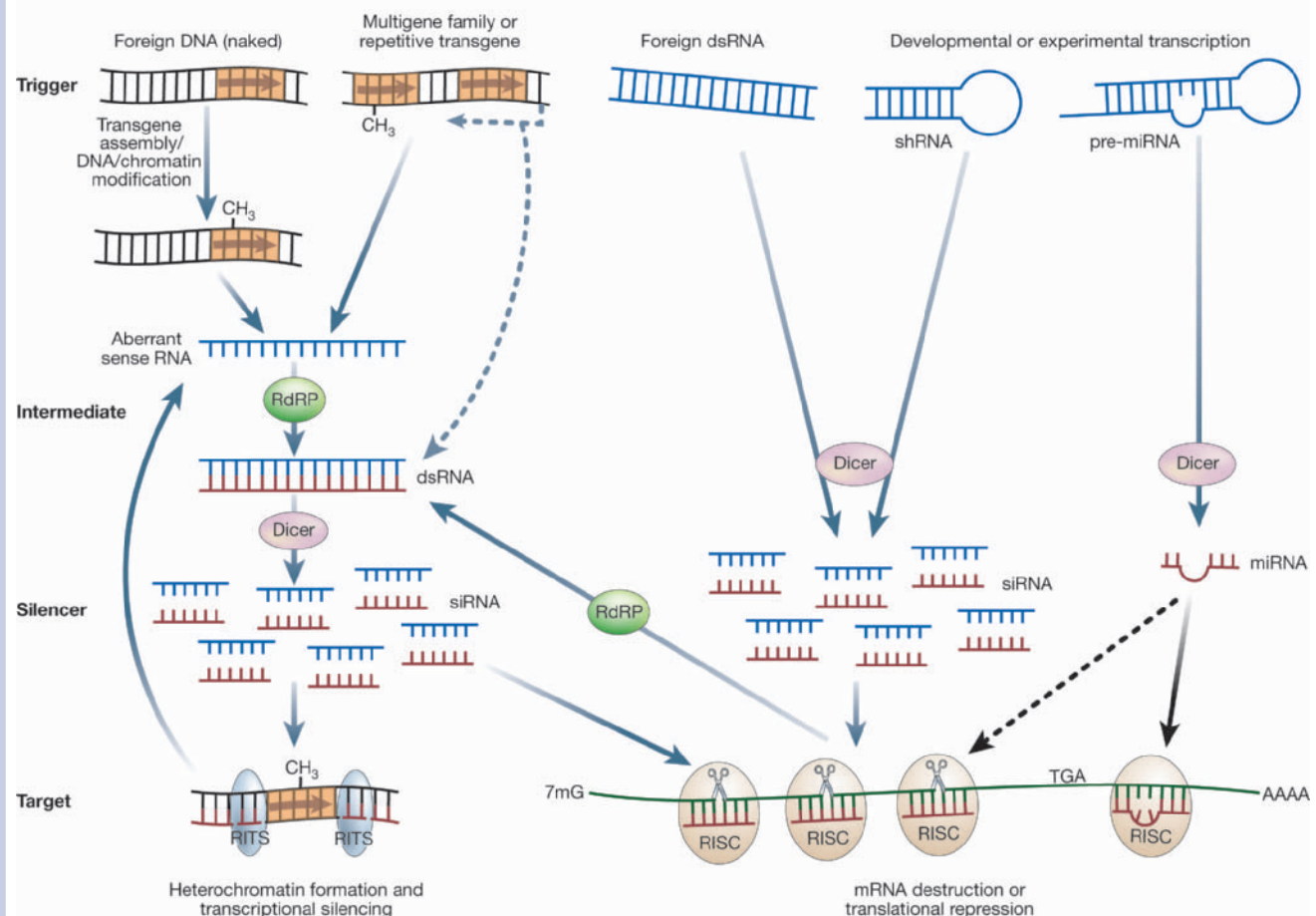


Figure 2 Model depicting distinct roles for dsRNA in a network of interacting silencing pathways. In some cases dsRNA functions as the initial stimulus (or trigger), for example when foreign dsRNA is introduced experimentally. In other cases dsRNA acts as an intermediate, for example when 'aberrant' mRNAs are copied by cellular RdRP. Transcription can produce dsRNA by readthrough from adjacent transcripts, as may occur for repetitive gene families or high-copy arrays (blue dashed arrows). Alternatively, transcription may be triggered experimentally or developmentally, for example in the expression of short hairpin (shRNA) genes and endogenous hairpin (miRNA) genes. The small RNA products of the Dicer-mediated dsRNA processing reaction guide distinct protein complexes to their targets. These silencing complexes include the RNA-induced silencing complex (RISC), which is implicated in mRNA destruction and translational repression, and the RNA-induced transcriptional silencing complex (RITS), which is implicated in chromatin silencing. Sequence mismatches between a miRNA and its target mRNA lead to translational repression (black solid arrow), whereas near perfect complementarity results in mRNA destruction (black dashed arrow). Feedback cycles permit an amplification and long-term maintenance of silencing. CH₃, modified DNA or chromatin; 7mG, 7-methylguanine; AAAA, poly-adenosine tail; TGA, translation termination codon.

injecting, feeding or expressing dsRNA¹⁴. However, RDE-1 and RDE-4 are not required for transposon silencing or for co-suppression^{14,50,51}. Furthermore, RDE-1 and RDE-4 are not required for the inheritance of RNAi-induced silencing⁹, which suggests that they are only required during the initial exposure to dsRNA. These findings indicate that transposon silencing and co-suppression in *C. elegans* are initiated by means of distinct triggers. As discussed above, an appealing idea is that a chromatin 'signature' stimulates the production of aberrant transcripts and the formation of a novel species of dsRNA (perhaps nuclear) that is distinct from the dsRNA that initiates silencing by means of RDE-1 and RDE-4. Again, in this model the initial trigger is the chromatin structure of the transposon locus or the transgene, and dsRNA acts as an intermediate in the silencing pathway (Fig. 2). Perhaps a similar RdRP-derived dsRNA functions in the RDE-1- and RDE-4-independent mechanisms that propagate silencing from one generation to the next.

Ten years ago, *de novo* cytosine methylation of genomic DNA was shown to occur in plants infected with RNA viroids whose sequences were homologous to the methylated genomic sequences⁵². This process was referred to as RNA-directed DNA methylation (RdDM).

Subsequently, dsRNA targeting a promoter was shown to trigger RdDM and initiate transcriptional silencing. The silencing was accompanied by the production of siRNAs⁵³, pointing to an RNAi-like mechanism for the initiation of transcriptional gene silencing. Recent work in fission yeast has now convincingly demonstrated that the formation of silent heterochromatin can be guided by small RNAs⁵⁴ and the RNA-silencing machinery⁵⁶. In *Drosophila*, the RNA-silencing machinery was also required for heterochromatin formation and for silencing multicopy transgenes and pericentric DNA⁵⁵. The discovery of an underlying molecular connection between RNA guides and chromatin remodelling has been one of the most exciting recent developments in the field of epigenetics. It is becoming clear that RNAi has an important role in the initiation of heterochromatin formation and transcriptional silencing in plants, fungi and animals (see review in this issue by Lippman and Martienssen, page 364).

The possibility of feedback between RNAi, its potential chromatin-associated trigger, and chromatin-mediated silencing maintenance mechanisms raises further questions about the ultimate causes of silencing. For example, were *C. elegans* transposons originally silenced by means of an RDE-1/RDE-4-dependent dsRNA signal, resulting

from sense and antisense readthrough transcription from insertion points in the genome^{56,57}? Perhaps over time this initial dsRNA-triggered silencing signal was replaced and augmented by a chromatin-associated silencing-maintenance signal.

Outlook for the RNA world

The numerous branching and converging silencing pathways that seem to exist in diverse organisms will no doubt require many years of research to unravel. It is already clear that different organisms have evolved distinct mechanisms, or at least variations on a common theme. In some cases, differences seem to exist in the extent to which silencing relies on a particular mode of regulation. For example, plants show a preponderance of miRNA-guided mRNA cleavage^{58,59}, but only one example of this mode of regulation has been found in animals⁶⁰. The diversity of RNA silencing phenomena suggests that other interesting findings await discovery. For example, the existence of an inheritance mechanism for the transmission of RNAi in *C. elegans* raises the question of whether natural small RNAs are transmitted in germ cells or other developmental cell lineages in other animals, including humans. Extrachromosomal inheritance of silencing patterns by means of small RNAs could provide sophisticated layers of gene regulation, at both post-transcriptional and chromatin-modifying levels. These small RNAs may be important in stem-cell maintenance and development, and differential localization of such RNAs may have a role in the generation of cellular diversity. It will be interesting to discover if the phenomenon of lateral transport of RNA from cell to cell, so far observed in plants^{61,62} and *C. elegans*, is more widespread. As well as having a role in immunity, could 'epigenetic RNA morphogens' allow cells to modulate the activity of developmentally important genes or mRNAs in neighbouring cells? This type of regulation might be particularly useful when cells, such as neurons, communicate at junctions that are far from the cell nucleus.

The past ten years have seen an explosion in the number of non-coding RNAs found to orchestrate remarkably diverse functions^{63,64}. These functions include: sequence-specific modification of cellular RNAs guided by small nucleolar RNAs⁶⁵; induction of chromosome-wide domains of chromatin condensation by the mammalian non-coding RNA Xist (X-inactive specific transcript)⁶⁶; autosomal gene imprinting and silencing by noncoding mammalian *Air* (antisense IgF2r RNA)⁶⁷; and finally sequence-directed cleavage and/or repression of target mRNAs and genes by miRNAs and siRNAs, discussed here and in the accompanying reviews. Some have likened this period to an RNA revolution. But considering the potential role of RNA as a primordial biopolymer of life, it is perhaps more apt to call it an RNA 'revelation'. RNA is not taking over the cell—it has been in control all along. We just didn't realize it until now. □

doi:10.1038/nature02872

- Joyce, G. F. & Orgel, L. E. In *The RNA World* (eds Gestland, R. F., Cech, T. R. & Atkins, J. F.) 49–77 (Cold Spring Harbor Laboratory Press, New York, 1999).
- Gilbert, W. The RNA world. *Nature* **319**, 618 (1986).
- Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
- Proud, C. G. PKR: a new name and new roles. *Trends Biochem. Sci.* **20**, 241–246 (1995).
- Williams, B. R. Role of the double-stranded RNA-activated protein kinase (PKR) in cell regulation. *Biochem. Soc. Trans.* **25**, 509–513 (1997).
- Guo, S. & Kemphues, K. J. *par-1*, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* **81**, 611–620 (1995).
- Nellen, W. & Lichtenstein, C. What makes an mRNA anti-sense-itiv? *Trends Biochem. Sci.* **18**, 419–423 (1993).
- Izant, J. G. & Weintraub, H. Inhibition of thymidine kinase gene expression by anti-sense RNA: a molecular approach to genetic analysis. *Cell* **36**, 1007–1015 (1984).
- Grishok, A., Tabara, H. & Mello, C. C. Genetic requirements for inheritance of RNAi in *C. elegans*. *Science* **287**, 2494–2497 (2000).
- Rochelleau, C. E. *et al.* Wnt signaling and an APC-related gene specify endoderm in early *C. elegans* embryos. *Cell* **90**, 707–716 (1997).
- Tabara, H., Grishok, A. & Mello, C. C. RNAi in *C. elegans*: soaking in the genome sequence. *Science* **282**, 430–431 (1998).
- Timmons, L. & Fire, A. Specific interference by ingested dsRNA. *Nature* **395**, 854 (1998).
- Timmons, L., Court, D. L. & Fire, A. Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene* **263**, 103–112 (2001).

- Tabara, H. *et al.* The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* **99**, 123–132 (1999).
- Schmidt, A. *et al.* Genetic and molecular characterization of *sting*, a gene involved in crystal formation and meiotic drive in the male germ line of *Drosophila melanogaster*. *Genetics* **151**, 749–760 (1999).
- Aravin, A. A. *et al.* Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr. Biol.* **11**, 1017–1027 (2001).
- Fagard, M., Boutet, S., Morel, J. B., Bellini, C. & Vaucheret, H. AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. *Proc. Natl Acad. Sci. USA* **97**, 11650–11654 (2000).
- Catalanotto, C., Azzalin, G., Macino, G. & Cogoni, C. Gene silencing in worms and fungi. *Nature* **404**, 245 (2000).
- de Carvalho, F. *et al.* Suppression of beta-1,3-glucanase transgene expression in homozygous plants. *EMBO J.* **11**, 2595–2602 (1992).
- Napoli, C., Lemieux, C. & Jorgensen, R. Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in *trans*. *Plant Cell* **2**, 279–289 (1990).
- Romano, N. & Macino, G. Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. *Mol. Microbiol.* **6**, 3343–3353 (1992).
- Kennerdell, J. R. & Carthew, R. W. Use of dsRNA-mediated genetic interference to demonstrate that *frizzled* and *frizzled 2* act in the wingless pathway. *Cell* **95**, 1017–1026 (1998).
- Ngo, H., Tschudi, C., Gull, K. & Ullu, E. Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proc. Natl Acad. Sci. USA* **95**, 14687–14692 (1998).
- Bosher, J. M. & Labouesse, M. RNA interference: genetic wand and genetic watchdog. *Nature Cell Biol.* **2**, E31–E36 (2000).
- Hamilton, A. J. & Baulcombe, D. C. A species of small antisense RNA in post-transcriptional gene silencing in plants. *Science* **286**, 950–952 (1999).
- Hammond, S. M., Bernstein, E., Beach, D. & Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* **404**, 293–296 (2000).
- Zamore, P. D., Tuschl, T., Sharp, P. A. & Bartel, D. P. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**, 25–33 (2000).
- Parrish, S. & Fire, A. Distinct roles for RDE-1 and RDE-4 during RNA interference in *Caenorhabditis elegans*. *RNA* **7**, 1397–1402 (2001).
- Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
- Elbashir, S. M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).
- Reinhart, B. J. *et al.* The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**, 901–906 (2000).
- Lee, R. C., Feinbaum, R. L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843–854 (1993).
- Hutvagner, G. *et al.* A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. *Science* **293**, 834–838 (2001).
- Ketting, R. F. *et al.* Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* **15**, 2654–2659 (2001).
- Grishok, A. A. *et al.* Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106**, 23–34 (2001).
- Volpe, T. A. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833–1837 (2002).
- Mochizuki, K., Fine, N. A., Fujisawa, T. & Gorovsky, M. A. Analysis of a *piwi*-related gene implicates small RNAs in genome rearrangement in *Tetrahymena*. *Cell* **110**, 689–699 (2002).
- Waterhouse, P. M., Graham, M. W. & Wang, M. B. Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc. Natl Acad. Sci. USA* **95**, 13959–13964 (1998).
- Dalmay, T., Hamilton, A., Rudd, S., Angell, S. & Baulcombe, D. C. An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for post-transcriptional gene silencing mediated by a transgene but not by a virus. *Cell* **101**, 543–553 (2000).
- Mourrain, P. *et al.* *Arabidopsis* SGS2 and SGS3 genes are required for post-transcriptional gene silencing and natural virus resistance. *Cell* **101**, 533–542 (2000).
- Cogoni, C. & Macino, G. Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. *Nature* **399**, 166–169 (1999).
- Sijen, T. *et al.* On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* **107**, 465–476 (2001).
- Smardon, A. *et al.* EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in *C. elegans*. *Curr. Biol.* **10**, 169–178 (2000).
- Schiebel, W., Haas, B., Marinkovic, S., Klanner, A. & Sanger, H. L. RNA-directed RNA polymerase from tomato leaves. II. Catalytic *in vitro* properties. *J. Biol. Chem.* **268**, 11858–11867 (1993).
- Makeyev, E. V. & Bamford, D. H. Cellular RNA-dependent RNA polymerase involved in post-transcriptional gene silencing has two distinct activity modes. *Mol. Cell* **10**, 1417–1427 (2002).
- Dougherty, W. G. & Parks, T. D. Transgenes and gene suppression: telling us something new? *Curr. Opin. Cell Biol.* **7**, 399–405 (1995).
- Baulcombe, D. C. RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants. *Plant Mol. Biol.* **32**, 79–88 (1996).
- Newport, J. Nuclear reconstitution *in vitro*: stages of assembly around protein-free DNA. *Cell* **48**, 205–217 (1987).
- Tabara, H., Yigit, E., Siomi, H. & Mello, C. C. The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in *C. elegans*. *Cell* **109**, 861–871 (2002).
- Dernburg, A. F., Zalevsky, J., Colaiacovo, M. P. & Villeneuve, A. M. Transgene-mediated co-suppression in the *C. elegans* germ line. *Genes Dev.* **14**, 1578–1583 (2000).
- Ketting, R. F. & Plasterk, R. H. A genetic link between co-suppression and RNA interference in *C. elegans*. *Nature* **404**, 296–298 (2000).
- Wassenegger, M., Heimes, S., Riedel, L. & Sanger, H. L. RNA-directed *de novo* methylation of genomic sequences in plants. *Cell* **76**, 567–576 (1994).
- Mette, M. F., van der Winden, J., Matzke, M. A. & Matzke, A. J. Production of aberrant promoter transcripts contributes to methylation and silencing of unlinked homologous promoters *in trans*. *EMBO J.* **18**, 241–248 (1999).

54. Reinhart, B. J. & Bartel, D. P. Small RNAs correspond to centromere heterochromatic repeats. *Science* **297**, 1831 (2002).
55. Pal-Bhadra, M. *et al.* Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery. *Science* **303**, 669–672 (2004).
56. Ketting, R. F., Haverkamp, T. H., van Luenen, H. G. & Plasterk, R. H. *mut-7* of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* **99**, 133–141 (1999).
57. Sijen, T. & Plasterk, R. Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature* **426**, 310–314 (2003).
58. Llave, C., Xie, Z., Kasschau, K. D. & Carrington, J. C. Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* **297**, 2053–2056 (2002).
59. Tang, G., Reinhart, B. J., Bartel, D. P. & Zamore, P. D. A biochemical framework for RNA silencing in plants. *Genes Dev.* **17**, 49–63 (2003).
60. Yekta, S., Shih, I. H. & Bartel, D. P. MicroRNA-directed cleavage of *HOXB8* mRNA. *Science* **304**, 594–596 (2004).
61. Palauqui, J. C., Elmayan, T., Pollien, J. M. & Vaucheret, H. Systemic acquired silencing: transgene-specific post-transcriptional silencing is transmitted by grafting from silenced stocks to non-silenced scions. *EMBO J.* **16**, 4738–4745 (1997).
62. Voinnet, O. & Baulcombe, D. C. Systemic signalling in gene silencing. *Nature* **389**, 553 (1997).
63. Storz, G. An expanding universe of noncoding RNAs. *Science* **296**, 1260–1263 (2002).
64. Morey, C. & Avner, P. Employment opportunities for non-coding RNAs. *FEBS Lett.* **567**, 27–34 (2004).
65. Kiss, T. Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell* **109**, 145–148 (2002).
66. Wutz, A. & Jaenisch, R. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol. Cell* **5**, 695–705 (2000).
67. Sleutels, F., Zwart, R. & Barlow, D. P. The non-coding *Air* RNA is required for silencing autosomal imprinted genes. *Nature* **415**, 810–813 (2002).

Acknowledgements C.C.M. is an HHMI Assistant Investigator and is funded by the NIH. D.C. is supported by an NRSA postdoctoral fellowship.

Competing interests statement The authors declare that they have no competing financial interests.

Mechanisms of gene silencing by double-stranded RNA

Gunter Meister & Thomas Tuschl

Laboratory of RNA Molecular Biology, The Rockefeller University, 1230 York Avenue, Box 186, New York, New York 10021, USA
(e-mail: ttuschl@rockefeller.edu)

Double-stranded RNA (dsRNA) is an important regulator of gene expression in many eukaryotes. It triggers different types of gene silencing that are collectively referred to as RNA silencing or RNA interference. A key step in known silencing pathways is the processing of dsRNAs into short RNA duplexes of characteristic size and structure. These short dsRNAs guide RNA silencing by specific and distinct mechanisms. Many components of the RNA silencing machinery still need to be identified and characterized, but a more complete understanding of the process is imminent.

First discovered in plants, where it was known as post-transcriptional gene silencing, RNA silencing or RNA interference (RNAi)¹ occurs in a wide variety of eukaryotic organisms^{2,3}. It is triggered by dsRNA precursors that vary in length and origin. These dsRNAs are rapidly processed into short RNA duplexes of 21 to 28 nucleotides in length, which then guide the recognition and ultimately the cleavage or translational repression of complementary single-stranded RNAs, such as messenger RNAs or viral genomic/antigenomic RNAs. The short RNAs have also been implicated in guiding chromatin modification (see review in this issue by Lippman and Martienssen, page 364).

According to their origin or function, three types of naturally occurring small RNA have been described: short interfering RNAs (siRNAs), repeat-associated short interfering RNAs (rasiRNAs) and microRNAs (miRNAs). In nature, dsRNA can be produced by RNA-templated RNA polymerization (for example, from viruses) or by hybridization of overlapping transcripts (for example, from repetitive sequences such as transgene arrays or transposons). Such dsRNAs give rise to siRNAs or rasiRNAs, which generally guide mRNA degradation and/or chromatin modification. In addition, endogenous transcripts that contain complementary or near-complementary 20- to 50-base-pair inverted repeats fold back on themselves to form dsRNA hairpins. These dsRNAs are processed into miRNAs that mediate translational repression, although they may also guide mRNA degradation. Finally, artificial introduction of long dsRNAs or siRNAs has been adopted as a tool to inactivate gene expression, both in cultured cells and in living organisms.

RNA silencing mechanisms were first recognized as antiviral mechanisms that protect organisms from RNA viruses⁴, or which prevent the random integration of transposable elements. But the general role of silencing in the regulation of gene expression only became apparent when it was realized that specific genes in plants and animals encode short forms of fold-back dsRNA⁵ (the precursor molecules of miRNAs; see review in this issue by Ambros, page 350). Many of these miRNA genes are evolutionarily conserved. In plants, miRNAs mainly function as siRNAs that guide the cleavage of sequence-complementary mRNAs. In animals such as the nematode *Caenorhabditis elegans* miRNAs appear predominantly to inhibit translation by targeting partially complementary sequences located within the 3' untranslated region (UTR) of mRNAs.

Here, we review the mechanisms of RNA gene silencing, and the roles of the proteins that make up the cellular post-transcriptional RNA silencing machinery. We discuss how dsRNAs are processed into small RNAs, how they are incorporated into effector complexes, and what the functions of these various complexes are. Finally, we discuss cellular regulatory processes and viral mechanisms that modulate RNA silencing efficiency. The picture that emerges is that RNA silencing is an evolutionarily conserved gene-regulatory mechanism with many species-specific variations, for example in the origin of the dsRNAs and in the number of homologous RNA silencing proteins expressed.

Processing dsRNA precursors

The maturation of small RNAs is a stepwise process catalysed by dsRNA-specific RNase-III-type endonucleases, termed Drosha and Dicer, which contain catalytic RNase III and dsRNA-binding domains (dsRBDs) (Figs 1 and 2). Drosha is specifically required for the processing of miRNA precursors, but not for the processing of long dsRNA. miRNAs are transcribed as long primary transcripts, which are first processed by Drosha in the nucleus^{6,7}. When Drosha excises the fold-back miRNA precursor, a 5' phosphate and a 2-nucleotide 3' overhang remain at the base of the stem^{7,8}. The miRNA precursor is then exported to the cytoplasm by means of the nuclear export receptor, exportin-5 (refs 9–11). Because exportin-5 lacks an obvious single-stranded or double-stranded RBD, it is not known whether the miRNA precursor binds directly to exportin-5 or to an RNA-binding adaptor protein.

Once it is in the cytoplasm, the miRNA precursor is further processed by Dicer^{12–16}. Processing of dsRNAs by Dicer yields RNA duplexes of about 21 nucleotides in length, which — like Drosha-processing products — have 5' phosphates and 2-nucleotide 3' overhangs¹⁷. Several organisms contain more than one Dicer gene, with each Dicer preferentially processing dsRNAs that come from a specific source (Fig. 1). *Drosophila melanogaster* has two paralogues: Dicer-1 (DCR-1) preferentially processes miRNA precursors¹⁴, and Dicer-2 (DCR-2) is required for long dsRNA processing^{14,18,19}. DCR-2 is stably associated with the dsRBD-containing protein R2D2 (ref. 18), and this complex then associates with siRNA duplexes (as monitored by gel-shift analysis¹⁹). The processing of dsRNA to siRNAs by the recombinant DCR-2 monomer or by the DCR-2/R2D2 heterodimer is ATP-dependent and requires a functional RNA helicase domain in DCR-2 (refs 14, 18, 20). For

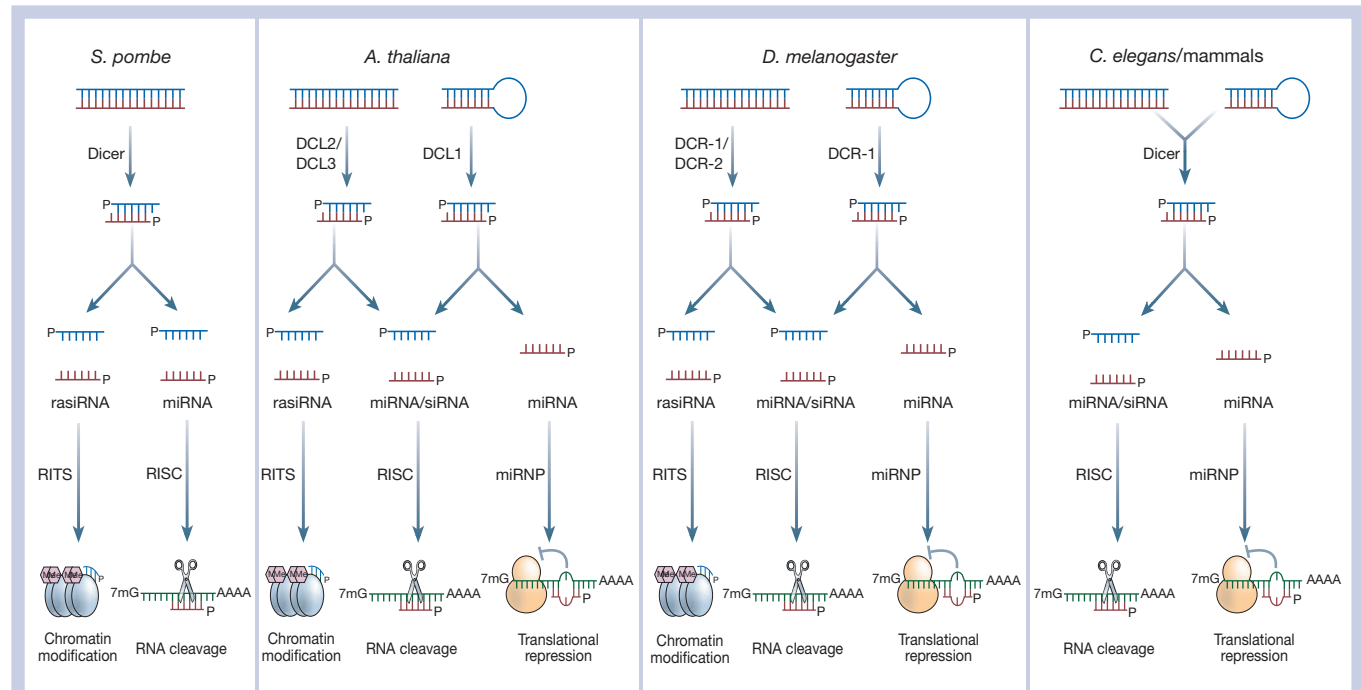


Figure 1 RNA silencing pathways in different organisms. Long dsRNA and miRNA precursors are processed to siRNA/miRNA duplexes by the RNase-III-like enzyme Dicer. The short dsRNAs are subsequently unwound and assembled into effector complexes: RISC, RITS (RNA-induced transcriptional silencing) or miRNP. RISC mediates mRNA-target degradation, miRNPs guide translational repression of target mRNAs, and the RITS complex guides the condensation of heterochromatin. In animals, siRNAs guide cleavage of complementary target RNAs, whereas miRNAs mediate translational repression of mRNA targets. rasiRNAs guide chromatin modification. *S. pombe*, *C. elegans* and mammals carry only one Dicer gene. In *D. melanogaster* and *A. thaliana*, specialized Dicer or DCL proteins preferentially process long dsRNA or miRNA precursors. 7mG, 7-methyl guanine; AAAAA, poly-adenosine tail; Me, methyl group; P, 5' phosphate.

human recombinant Dicer, however, ATP-dependence of dsRNA processing has not been observed^{21,22}. The reason for this species difference in ATP-dependence is unclear.

Four Dicer-like (DCL) proteins (DCL1 to DCL4) have been identified in *Arabidopsis thaliana*, three of which are involved in processing dsRNAs that come from different sources¹⁵. DCL1 processes miRNA precursors¹⁵, requiring two more proteins to do so: HEN1 (ref. 23) and the dsRBD protein HYL1 (refs 24, 25). DCL2 is required for the production of siRNAs from plant viruses, and DCL3, which also cooperates with HEN1 (ref. 15), is involved in the production of rasiRNAs. In *C. elegans*, only one Dicer (DCR-1) has been identified. This cooperates with the dsRBD protein RDE-4 during RNAi, although RDE-4 is not required for miRNA function²⁶.

It is conceivable that in *C. elegans* and mammals, which possess only a single Dicer gene, as-yet-unidentified Dicer-interacting dsRBD-containing proteins allow Dicer to recognize different sources of dsRNA.

Assembly into RNA silencing effector complexes

The siRNA- and miRNA-duplex-containing ribonucleoprotein particles (RNPs) are subsequently rearranged into the RNA-induced silencing complex (RISC)²⁷. The functional RNPs contain only single-stranded siRNAs or miRNAs. Although it is difficult to assign distinct functional labels at this point, an siRNA-containing effector complex is commonly referred to as a RISC, whereas a miRNA-containing effector complex is referred to as a miRNP²⁸. Every RISC or miRNP contains a member of the Argonaute (Ago) protein family; the Ago protein probably binds directly to the RNA in these complexes, although formal evidence for this is lacking^{28–31}. Several forms of RISC or miRNPs that differ in size and composition have been reported, and these presumably differ in activity or even in function. The estimated apparent molecular mass ranges from between 130 to 160 kDa in the case of high-salt purified RISC from human cells^{30,32} to 500 kDa^{20,29} or up to the 80S range in the case of

RISC isolated from *D. melanogaster* cell lysates¹⁹. The differences in mass are unlikely to be caused by oligomerization of smaller RISC units (minimal RISC). Instead, the gain in molecular mass appears to be due to the weak and/or transient association of proteins involved in the initial processing of dsRNA (Dicer, R2D2)^{19,29,33}, and of other factors of unknown function.

The assembly of RISC and presumably also of miRNPs is ATP-dependent^{19,20}, which probably reflects the requirement for energy-driven unwinding of the siRNA- or miRNA-duplex and/or other conformational or compositional changes of the pre-assembled RNA-duplex-containing RNP. Likely candidates for factors that promote these ATP-dependent conformational changes are DEAD-box RNA helicases (Fig. 3). Several ATPases have been implicated in RNA silencing (Tables 1 and 2), but only one has been characterized in detail. In *D. melanogaster*, the putative DEAD-box RNA helicase Armitage is required after Dicer processing for RISC assembly³⁴ and presumably also for miRNP assembly³⁵.

In the course of identifying more active and more specific siRNA duplexes to guide mRNA cleavage, it was noticed that the sequence composition of the siRNA duplex has an impact on the ratio of 'sense' (same sequence as the target gene) and 'antisense' (complementary to the target gene) siRNAs entering the RISC complex^{36,37}. Naturally occurring miRNAs also show a strong bias for accumulating only one strand into a miRNP⁵. Effective, or more potently silencing, siRNAs or miRNA duplexes show reduced thermodynamic stability at the 5' end of the antisense siRNA or miRNA relative to the 3' end within the duplex. This strand bias is presumably caused by a rate-limiting unwinding step that occurs during the transition from the duplex-containing RNP to the larger RISC/miRNP complex, which allows the 5' end of the strand positioned at the weakly paired end of the dsRNA to enter RISC/miRNP first.

The single-stranded siRNAs/miRNAs residing in RISC/miRNP are extremely tightly bound to an Ago protein: salt concentrations as high as 2.5 M KCl do not affect the association of the small RNA with

the Ago protein during affinity purification of RISC³². Ago proteins have a molecular mass of about 100 kDa and contain two conserved domains: PAZ (for piwi–argonaute–zwillie) and PIWI (Fig. 3)³⁸. The PIWI domain has been implicated in an interaction with Dicer³³, although the analysis was carried out without using the full-length proteins. Recently, the crystal structure of an archaeobacterial Ago protein revealed striking similarity of the PIWI domain with members of the RNase H family³⁹. As RNase H cleaves the RNA strand of RNA/DNA duplexes, it was proposed that Ago proteins act by cleaving target RNA in target RNA/siRNA hybrids. Initially, the PAZ domain was suggested to function as a protein–interaction domain between Ago and Dicer, because a PAZ domain is also present in many Dicer–RNase-III enzymes and Ago proteins co-immunoprecipitate with Dicer²⁹. Recent biochemical and structural studies, however, converged on the view that PAZ is an RBD that specifically recognizes the terminus of the base-paired helix of siRNA and miRNA duplexes, including the characteristic 2-nucleotide 3' overhangs^{40,41}. This siRNA/miRNA-duplex-specific interaction with PAZ ensures the safe transition of small RNAs into RISC/miRNP by minimizing the possibility of unrelated RNA-processing or RNA-turnover products entering the RNA silencing pathway. The preferred recognition of the termini of dsRNA²² or miRNA precursors⁸ by PAZ-domain-containing Dicer suggests that the processing reaction is guided by the PAZ domain docking at the terminus of long dsRNAs.

A possible variation in the general mechanism of siRNA transfer from the dsRNA-processing complex into RISC has been observed for Dicer variants that presumably lack the PAZ domain¹⁸. For example, *D. melanogaster* DCR-2, which predominantly processes siRNAs, requires R2D2 — not for dsRNA processing but for the incorporation of siRNA into RISC¹⁸. R2D2 contains two dsRBDs. The binding mode of DCR-2 and R2D2 may expose the 2-nucleotide staggered terminus of the processed siRNA for docking of the Ago protein and its PAZ domain. The 2-nucleotide 3' overhanging structure of siRNA duplexes is essential for effective assembly of RISC in *D. melanogaster* lysate⁴². This implies that a step exists during RISC assembly that involves specific recognition of the 2-nucleotide overhang, presumably by the incoming Ago protein and its PAZ domain. The *C. elegans* RDE-4 protein, which is structurally related to R2D2, was shown to bind to the RDE-1 Ago protein²⁶, and therefore may have a similar function to R2D2 in forming active RISC.

The way in which single-stranded siRNA/miRNA binds to Ago after unwinding of the small RNA duplex is not understood because recombinant full-length Ago proteins are difficult to express and therefore the reconstitution of a functional complex containing a defined RNA sequence has not yet been accomplished.

Different organisms have different numbers of Ago proteins, ranging from one in *Schizosaccharomyces pombe* to more than 20 in *C. elegans*³⁸. In *A. thaliana* ten members have been identified⁴³, compared with five in *D. melanogaster*⁴⁴ and eight in humans⁴⁵. So far, only a small subset of this family has been functionally characterized; the specific functions of these Ago proteins are summarized in Tables 1 and 2. The evidence suggests that the different Ago proteins are not redundant. In species expressing more than one Dicer protein, the specificity of small-RNA loading into RISC/miRNP is probably controlled by the individual Dicer and Ago interactions. In species with a single Dicer, it is unclear whether and how the loading of the different Ago protein members is controlled for the different sources of dsRNA. Ago proteins may have evolved an intrinsic sequence-specificity that allows them to bind preferentially to small RNAs of specific sequence, or they may use specific adaptor proteins that are associated with the different cellular sources of dsRNA. A recent study in *D. melanogaster* revealed that AGO1 is required for miRNA accumulation whereas AGO2 is required for siRNA-triggered mRNA degradation⁴⁶. Moreover, two recent studies in human systems showed that AGO1–AGO4 all associate with miRNAs and siRNAs, but that only the AGO2-containing RNPs exhibit RISC activity^{47,48}. Finally, different expression patterns and levels of the various Ago proteins may control

Table 1 Factors involved in RNAi or transcriptional gene silencing

Protein	Domains and motifs	Function	Species	References
Dicer and Dicer-like proteins				
Dicer	RNase III, DEAD, PAZ, dsRBD	Long dsRNA processing	Hs	16
DCR-1	RNase III, DEAD, PAZ, dsRBD	Long dsRNA processing	Ce	13
DCR-1	RNase III, HELICc, PAZ, dsRBD	Long dsRNA processing, RISC assembly	Dm	14, 19
DCR-2	RNase III, DEAD, dsRBD	Long dsRNA processing, RISC assembly	Dm	14, 18, 19
DCL2	RNase III, DEAD, PAZ, dsRBD	Long dsRNA processing	At	15
DCL3	RNase III, DEAD, PAZ	Long dsRNA processing	At	15
Ago protein family				
AGO1	PAZ, PIWI	Short RNA binding	Hs, Dm	30, 44
AGO1	PAZ, PIWI		At, Sp	80, 81
AGO2	PAZ, PIWI	Short RNA binding	Hs, Dm	29, 30
AGO3	PAZ, PIWI	Short RNA binding	Hs	48
AGO4	PAZ, PIWI		At	82
Plwi	PAZ, PIWI		Dm	83
Aubergine	PAZ, PIWI		Dm	84
RDE-1	PAZ, PIWI	Initiation of RNAi	Ce	26
PPW-1	PAZ, PIWI		Ce	85
QDE-2	PAZ, PIWI		Nc	86
Putative RNA helicases				
Armitage	DEAD	RISC assembly	Dm	34
Spindle E	DEAD		Dm	84
Rm62	DEAD		Dm	51
SDE3	DEAD		At	87
DRH-1/2	DEAD		Ce	26
MUT-14	DEAD		Ce	88
SMG-2	DEAD		Ce	89
RNA-dependent RNA polymerases				
SGS2/SDE1	RNA polymerase	RNA amplification	At	90
RDR2	RNA polymerase	RNA amplification	At	91
EGO-1	RNA polymerase	RNA amplification	Ce	92
RRF-1	RNA polymerase	RNA amplification	Ce	67
RRF-3	RNA polymerase		Ce	74
QDE-1	RNA polymerase	RNA amplification	Nc	93
Other factors				
R2D2	dsRBD	RISC assembly	Dm	18
FMRp	KH, RGG	RNA binding, translational regulation	Dm	49
Vasa intronic gene (VIG)	RGG		Dm	49
Tudor-SN (p100)	Tudor, nuclease		Dm, Hs	50
WEX-1	Exonuclease		At	94
SGS3			At	90
SDE4			At	82
RDE-4	dsRBD	Initiation of RNAi	Ce	26
MUT-7	Exonuclease		Ce	95
ADAR-1/2	Deaminase	Adenosine deamination	Ce	71
SID-1		dsRNA transport	Ce	96
ERI-1	Exonuclease	siRNA degradation	Ce	73
QDE-3	DNA helicase		Nc	97
CHP1	Chromo	Heterochromatin association	Sp	81
TAS3			Sp	81

Hs, *Homo sapiens*; Dm, *D. melanogaster*; At, *A. thaliana*; Ce, *C. elegans*; Nc, *Neurospora crassa*; Sp, *S. pombe*.

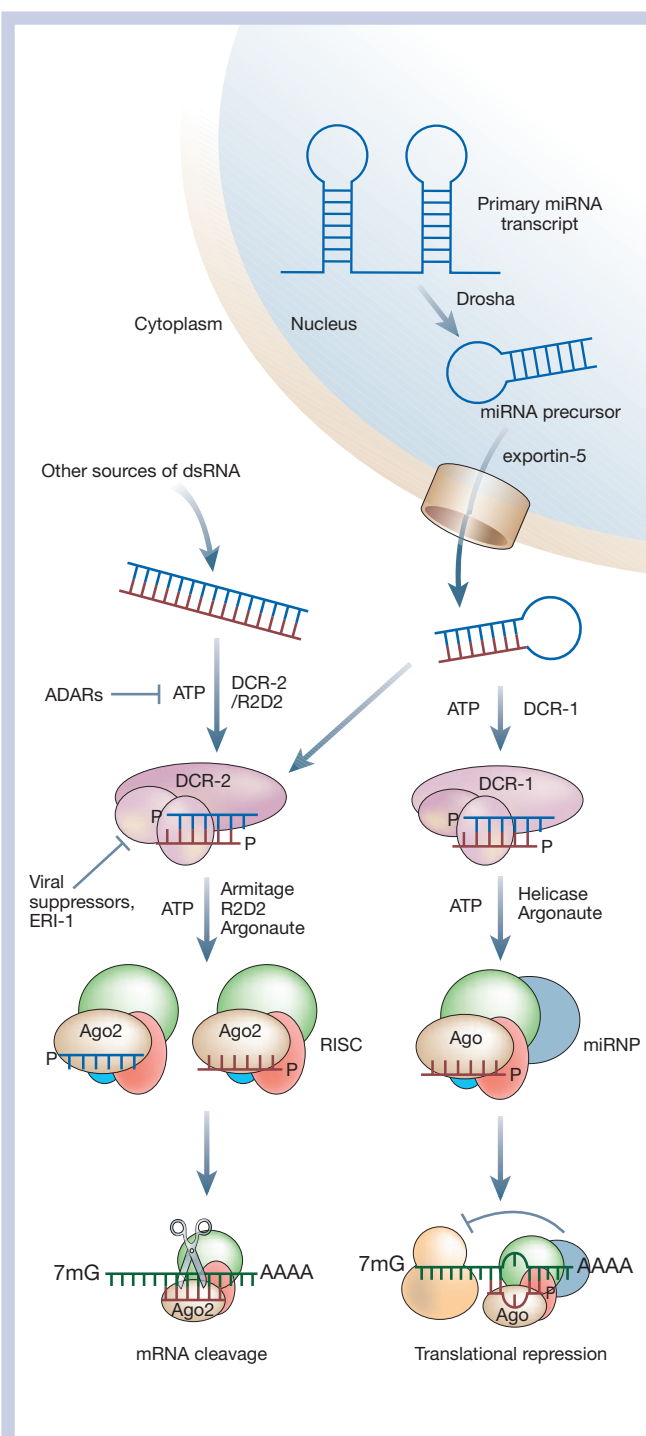


Figure 2 Model of small-RNA-guided post-transcriptional regulation of gene expression. Primary miRNA transcripts are processed to miRNA precursors in the nucleus by the RNase-III-like enzyme Drosha. The miRNA precursor is subsequently exported to the cytoplasm by means of the export receptor exportin-5. The miRNA precursor is further processed by Dicer to siRNA-duplex-like intermediates. The duplex is unwound while assembling into miRNP/RISC. Mature miRNAs bind to Ago proteins, which mediate translational repression or cleavage of target mRNAs. Other sources of long dsRNA in the cytoplasm of a cell are viral RNAs, artificially introduced dsRNA, dsRNAs generated by RdRPs, and genomic sense and antisense transcripts. Like miRNA precursors, long dsRNA is processed by the RNase III enzyme Dicer into 21–23 nucleotide dsRNA intermediates. Assisted by the RNA helicase Armitage and R2D2, the single-stranded siRNA-containing RISC is formed. The stability of the dsRNA and its recognition by Dicer can be regulated by specific ADARs and the exonuclease ERI-1.

the extent to which the different RNA silencing processes operate. The functional characterization of the individual Ago proteins remains a considerable challenge in the field.

Besides the core Ago protein and the small RNA component of the various RNA silencing complexes, other proteins have been identified in the larger forms of purified complexes. Three more components were identified in purified *D. melanogaster* RISC: the Vasa intronic gene product (VIG), fragile-X-related protein (dFXR), and the tudor staphylococcal-nuclease-domain-containing protein (Tudor-SN)^{49,50}. In a preparation of *D. melanogaster* miRNPs, AGO2, dFXR, the DEAD-box helicase RM62, and the ribosomal proteins RPL5 and RPL11 were identified, indicating a role of miRNPs in the regulation of mRNA translation⁵¹. In human biochemical systems, interactions of AGO2/eIF2C2 (eukaryotic translation initiation factor 2C2) with FMRP (fragile X mental retardation protein) — the human homologue of dFXR — has also been observed⁵². Furthermore, human miRNAs, which reside in a defined 15S complex, were shown to contain AGO2 and the RNA helicases Gemin3 and Gemin4 (ref. 28). The precise function of these different proteins in the RNA silencing process remains unclear.

mRNA cleavage and translational repression

The single-stranded siRNA in RISC guides sequence-specific degradation of complementary or near-complementary target mRNAs^{30,32}. RISC cleaves the target mRNA in the middle of the complementary region, ten nucleotides upstream of the nucleotide paired with the 5' end of the guide siRNA¹⁷. The cleavage reaction guided by RISC/miRNP does not require ATP^{20,31}. However, multiple rounds of mRNA cleavage — which requires the release of cleaved mRNA products — are more efficient in the presence of ATP³¹. ATP-dependent enhancement of multiple rounds of mRNA cleavage may reflect the assistance of an RNA helicase. This may facilitate the release of mRNA products from the RISC/miRNP complex, or reorganize the endonuclease complex for another round of target-RNA recognition.

RISC/miRNP complexes catalyse hydrolysis of the target-RNA phosphodiester linkage, yielding 5' phosphate and 3' hydroxyl termini^{32,53}. This hydrolysis reaction is different from classic pancreatic-RNase endonucleolytic cleavage, which occurs by nucleophilic attack of the adjacent 2' hydroxyl group and yields a 2',3'-cyclic phosphate intermediate. The target-RNA hydrolysis reaction requires magnesium ions and resembles mechanistically the hydrolysis reaction that occurs when Dicer generates siRNA duplexes from dsRNA precursors — a reaction that also requires magnesium ions^{21,22}. However, the small molecular mass of minimal RISC argues against a role for Dicer in directly catalysing target-RNA cleavage^{30,32}. The Tudor-SN nuclease, which has been characterized as a component of RISC⁵⁰, can also be eliminated as a candidate for the unidentified endonuclease activity in RISC because, as a member of the micrococcal nuclease protein family, it should produce 3' phosphate rather than 5' phosphate termini. A smaller double-strand-specific endonuclease may be part of RISC, or as has recently been proposed on the basis of the similarity of the PIWI domain with RNase H⁴⁰, Ago proteins themselves may act as nucleases. However, it remains unresolved, why nuclease activity has been detected with Ago2 but not Ago1 or Ago3 in human cells, even though the proposed catalytic DDE motif is intact in all three proteins. It has been speculated that the RNA subunit of RISC may assist catalysis, but this is unlikely given that base pairing during target recognition positions the small RNA sugar-phosphate backbone far away from the active site. In animals, the binding sites for miRNAs in target mRNAs are generally considered to be of insufficient complementarity to allow the target to be cleaved. Recently, however, it was found that RISC/miRNP can cleave significantly mismatched targets, albeit at a reduced rate, although this depends on the type, position and structure of the pairing interaction³².

The mechanism of miRNA-guided translational regulation is not as well understood as that of target-RNA cleavage. The first evidence for

Table 2 Factors involved in miRNA guided translational regulation

Protein	Motifs	Function	Species	References
Dicer and Dicer-like proteins				
Dicer	RNase III, DEAD, PAZ, dsRBD	miRNA precursor processing	Hs	12
DCR-1	RNase III, DEAD, PAZ, dsRBD	miRNA precursor processing	Ce	13
DCR-1	RNase III, HELICc, PAZ, dsRBD	miRNA precursor processing	Dm	14
DCL1	RNase III, DEAD, PAZ, dsRBD	miRNA precursor processing	At	15
Ago protein family				
AGO1	PAZ, PIWI	Short RNA binding	Dm, Hs	46
AGO2	PAZ, PIWI	Short RNA binding	Dm, Hs	28, 51, 52
AGO3	PAZ, PIWI	Short RNA binding	Hs	47
AGO4	PAZ, PIWI	Short RNA binding	Hs	47
AGO1	PAZ, PIWI		At	98
ALG-1/2	PAZ, PIWI		Ce	13
Putative RNA helicases				
Gemin3	DEAD	RNA helicase	Hs	28
Armitage	DEAD	RISC assembly	Dm	35
Other factors				
Drosha	RNase III	Processing of primary miRNA transcripts	Hs	7, 8
Exportin-5		Nuclear export of miRNA precursors	Hs	9–11
Gemin4			Hs	28
FMRP, FXR	KH, RGG	RNA binding	Dm, Hs	51, 52
Ribosomal proteins L8, L11		60S ribosomal subunit	Dm	51
HYL1	dsRBD		At	24, 25
HEN1			At	23

translational repression by miRNAs was obtained from studies of mutant or transgenic *C. elegans*, where it was shown that miRNAs targeted to a specific gene reduced protein synthesis without affecting mRNA levels⁵. The target mRNA contained in its 3' UTR several binding sites for the miRNA, and both target and miRNA were found to be associated with polyribosomes. This suggested that miRNAs block translation elongation or termination rather than translational initiation^{54,55}. Ribosome profiles obtained from mammalian cell extracts also indicate an association of a small fraction of miRNAs with polysomes⁵⁶.

To complicate matters, miRNAs can act as siRNAs³¹, and siRNAs can act as miRNAs^{57,58}. This observation has occasionally been over-interpreted as suggesting that the function of a siRNA or a miRNA is solely determined by the degree of complementarity between the small RNA and its target RNA. An alternative, and probably more accurate, interpretation is that small dsRNAs are assembled into characteristic complexes that carry out distinct functions, depending on the protein factors that they recruit. This model could partly explain the apparent redundancy of the Ago protein family. Presumably, miRNA-guided translational regulation and targeted mRNA degradation are used simultaneously as natural regulatory mechanisms.

Regulators of RNAi

RNA silencing controls downstream-target gene expression but is itself under regulation. For example, systemic silencing, caused by the spread of a silencing signal between cells, occurs in plants and nematodes, and requires a group of enzymes termed RNA-dependent RNA polymerases (RdRPs; also known as RDRs). RdRPs are thought to amplify the silencing effect by dsRNA synthesis of the target mRNAs and/or its cleavage products^{59,60}. In *D. melanogaster* and vertebrates, there are no equivalent RdRP-like proteins, and evidence exists against there being any amplification mechanism for silencing triggers^{61–63}.

In many organisms, RNA silencing mediated by dsRNA is part of an innate immune response against RNA viruses and transposable



Figure 3 Domain structures of representative members of the RNA family discussed in the review. Dicer and Ago proteins contain PAZ domains. Ago proteins carry an extra PIWI domain. Armitage and Dicer contain RNA-helicase domains. Dicer and Drosha both contain RNase III domains (RIIla, RIIlb). dsRBDs are present in Dicer, Drosha and R2D2. *Drosophila* DCR-1 and human Dicer contain a similar domain of unknown function (DUF283).

elements^{64,65}. Defence strategies to thwart the host response have been found in plant viruses^{66,67} and the insect flock house virus⁶⁸. These viruses express inhibitors, such as dsRNA-binding proteins, that interfere with the host cell's RNA silencing machinery. One of the viral suppressor proteins, p19, from the plant tombusvirus, has been co-crystallized with a duplex of 21-nucleotide siRNAs^{69,70}. A homodimer of p19 binds to the RNA sugar-phosphate backbone, and positions two α -helices on top of the terminal base pairs of the RNA duplex. This binding allows the p19 protein to discriminate siRNA duplexes from shorter or longer dsRNAs. Other inhibitors of RNA silencing in plants target RISC or restrict the spreading of a systemic silencing signal from infected cells^{66,67}. The precise chemical nature of this spreading silencing signal is still debated: dsRNA, siRNA, RISC, or as-yet-unidentified RNAs or RNPs are possible candidates.

dsRNA acts as a substrate from which Dicer can produce siRNAs, but it also competes with the 'RNA-editing' process catalysed by dsRNA-specific adenosine deaminases (ADARs), which convert adenosine to inosine. This editing reaction not only eliminates the complementarity between the dsRNA and the target mRNA, but also destabilizes the edited dsRNA, making it a poor substrate for Dicer^{71,72}. Another negative regulator of RNA silencing, the exonuclease-domain-containing protein ERI-1, was discovered in genetic screens in *C. elegans*⁷³. ERI-1 is specifically expressed in neurons, and loss-of-function *eri-1* mutants show increased RNAi in neurons. Biochemical analysis suggests that siRNA is a preferred substrate for ERI-1. The presumably inactive RdRP-family protein RRF-3 (ref. 74) is also a negative regulator of RNAi in *C. elegans*. RRF-3 presumably competes for templates or primers of RNA amplification, inhibiting the production of additional short RNAs.

Outlook

Understanding the mechanisms of RNA silencing may shed light on human disease. Recently, several links between RNA-silencing factors and inherited or acquired genetic disorders have been recognized. For example, the loss or mutation of FMRP that causes the neurological disorder, fragile X syndrome, indirectly affects the translational regulation of miRNA-targeted mRNAs that are also targeted by FMRP^{51,52}. Other studies have linked diseases to the loss of miRNA expression that may result from the deletion or the translocation of a

chromosomal region. The miRNAs miR-15 and miR-16 were identified in the tumour suppressor locus for B-cell chronic lymphocytic leukaemia (*B-CLL*)⁷⁵, and consistent down-regulation of miR-143 and miR-145 has been observed in advancing colorectal neoplasia⁷⁶. Translocation of the proto-oncogene, *c-myc*, downstream of the very strong promoter of miR-122 (ref. 77) and miR-142 (ref. 78) was also found in two cancer instances. Furthermore, it was recently discovered that some viruses express miRNAs, which presumably regulate viral as well as host gene expression⁷⁹. To understand the underlying specificity of these processes, we need to identify the regulated targets of miRNAs, using computational or experimental methods. The development of robust biochemical and other methods for the dissection of the function and mechanism of RNA silencing are likely to provide many more insights into RNA silencing machineries. □

doi:10.1038/nature02873

1. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
2. Tijsterman, M., Ketting, R. F. & Plasterk, R. H. The genetics of RNA silencing. *Annu. Rev. Genet.* **36**, 489–519 (2002).
3. Ullu, E., Tschudi, C. & Chakraborty, T. RNA interference in protozoan parasites. *Cell Microbiol.* **6**, 509–519 (2004).
4. Waterhouse, P. M., Wang, M. B. & Lough, T. Gene silencing as an adaptive defence against viruses. *Nature* **411**, 834–842 (2001).
5. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
6. Lee, Y., Jeon, K., Lee, J. T., Kim, S. & Kim, V. N. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J.* **21**, 4663–4670 (2002).
7. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).
8. Basyuk, E., Suavet, F., Doglio, A., Bordonne, R. & Bertrand, E. Human let-7 stem-loop precursors harbor features of RNase III cleavage products. *Nucleic Acids Res.* **31**, 6593–6597 (2003).
9. Bohnsack, M. T., Czaplinski, K. & Görlich, D. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* **10**, 185–191 (2004).
10. Lund, E., Guttlinger, S., Calado, A., Dahlberg, J. E. & Kutay, U. Nuclear export of microRNA precursors. *Science* **303**, 95–98 (2004).
11. Yi, R., Qin, Y., Macara, I. G. & Cullen, B. R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* **17**, 3011–3016 (2003).
12. Hutvagner, G., McLachlan, J., Balint, E., Tuschl, T. & Zamore, P. D. A cellular function for the RNA interference enzyme Dicer in small temporal RNA maturation. *Science* **93**, 834–838 (2001).
13. Grishok, A. *et al.* Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106**, 23–34 (2001).
14. Lee, Y. S. *et al.* Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**, 69–81 (2004).
15. Xie, Z. *et al.* Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2**, E104 (2004).
16. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
17. Elbashir, S. M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21 and 22 nt RNAs. *Genes Dev.* **15**, 188–200 (2001).
18. Liu, Q. *et al.* R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science* **301**, 1921–1925 (2003).
19. Pham, J. W., Pellino, J. L., Lee, Y. S., Carthew, R. W. & Sontheimer, E. J. A Dicer-2-dependent 80S complex cleaves targeted mRNAs during RNAi in *Drosophila*. *Cell* **117**, 83–94 (2004).
20. Nykänen, A., Haley, B. & Zamore, P. D. ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell* **107**, 309–321 (2001).
21. Provost, P. *et al.* Ribonuclease activity and RNA binding of recombinant human Dicer. *EMBO J.* **21**, 5864–5874 (2002).
22. Zhang, H., Kolb, F. A., Brondani, V., Billy, E. & Filipowicz, W. Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J.* **21**, 5875–5885 (2002).
23. Boutet, S. *et al.* *Arabidopsis* HEN1: A genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance. *Curr. Biol.* **13**, 843–848 (2003).
24. Vazquez, F., Gascioli, V., Crete, P. & Vaucheret, H. The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not post-transcriptional transgene silencing. *Curr. Biol.* **14**, 346–351 (2004).
25. Han, M. H., Goud, S., Song, L. & Fedoroff, N. The *Arabidopsis* double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proc. Natl Acad. Sci. USA* **101**, 1093–1098 (2004).
26. Tabara, H., Yigit, E., Siomi, H. & Mello, C. C. The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DEXH-box helicase to direct RNAi in *C. elegans*. *Cell* **109**, 861–871 (2002).
27. Hammond, S. M., Bernstein, E., Beach, D. & Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* **404**, 293–296 (2000).
28. Mourelatos, Z. *et al.* miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.* **16**, 720–728 (2002).
29. Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R. & Hannon, G. J. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**, 1146–1150 (2001).
30. Martinez, J., Patkaniowska, A., Urlaub, H., Lührmann, R. & Tuschl, T. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**, 563–574 (2002).
31. Hutvagner, G. & Zamore, P. D. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**, 2056–2060 (2002).
32. Martinez, J. & Tuschl, T. RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes Dev.* **18**, 975–980 (2004).
33. Tabbaz, N. *et al.* Characterization of the interactions between mammalian PAZ PIWI domain proteins and Dicer. *EMBO Rep.* **5**, 189–194 (2004).
34. Tomari, Y. *et al.* RISC assembly defects in the *Drosophila* RNAi mutant *armitage*. *Cell* **116**, 831–841 (2004).
35. Cook, H. A., Koppetsch, B. S., Wu, J. & Theurkauf, W. E. The *Drosophila* SDE3 homolog *armitage* is required for *oskar* mRNA silencing and embryonic axis specification. *Cell* **116**, 817–829 (2004).
36. Khvorovova, A., Reynolds, A. & Jayasena, S. D. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209–216 (2003).
37. Schwarz, D. S. *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199–208 (2003).
38. Carmell, M. A., Xuan, Z., Zhang, M. Q. & Hannon, G. J. The Argonaute family: tentacles that reach into RNAi, developmental control, stem-cell maintenance, and tumorigenesis. *Genes Dev.* **16**, 2733–2742 (2002).
39. Song, J. J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science*; published online 29 July 2004 (doi:10.1126/science.1102514).
40. Ma, J. B., Ye, K. & Patel, D. J. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **429**, 318–322 (2004).
41. Lingel, A., Simon, B., Izaurralde, E. & Sattler, M. Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nature Struct. Mol. Biol.* **11**, 576–577 (2004).
42. Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W. & Tuschl, T. Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.* **20**, 6877–6888 (2001).
43. Hunter, C., Sun, H. & Poethig, R. S. The *Arabidopsis* heterochronic gene *ZIPPY* is an ARGONAUTE family member. *Curr. Biol.* **13**, 1734–1739 (2003).
44. Williams, R. W. & Rubin, G. M. ARGONAUTE1 is required for efficient RNA interference in *Drosophila* embryos. *Proc. Natl Acad. Sci. USA* **99**, 6889–6894 (2002).
45. Sasaki, T., Shiohama, A., Minoshima, S. & Shimizu, N. Identification of eight members of the Argonaute family in the human genome. *Genomics* **82**, 323–330 (2003).
46. Okamura, K., Ishizuka, A., Siomi, H. & Siomi, M. C. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* **18**, 1655–1666 (2004).
47. Meister, G. *et al.* Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell* **15**, 185–197 (2004).
48. Liu, J. *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* published online 29 July 2004 (doi: 10.1126/science.1102513).
49. Caudy, A. A., Myers, M., Hannon, G. J. & Hammond, S. M. Fragile X-related protein and VIG associate with the RNA interference machinery. *Genes Dev.* **16**, 2491–2496 (2002).
50. Caudy, A. A. *et al.* A micrococcal nuclease homologue in RNAi effector complexes. *Nature* **425**, 411–414 (2003).
51. Ishizuka, A., Siomi, M. C. & Siomi, H. A *Drosophila* fragile X protein interacts with components of RNAi and ribosomal proteins. *Genes Dev.* **16**, 2497–2508 (2002).
52. Jin, P. *et al.* Biochemical and genetic interaction between the fragile X mental retardation protein and the microRNA pathway. *Nature Neurosci.* **7**, 113–117 (2004).
53. Schwarz, D. S., Tomari, Y. & Zamore, P. D. The RNA-induced silencing complex is a Mg²⁺-dependent endonuclease. *Curr. Biol.* **14**, 787–791 (2004).
54. Olsen, P. H. & Ambros, V. The *lin-4* regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev. Biol.* **216**, 671–680 (1999).
55. Seggerson, K., Tang, L. & Moss, E. G. Two genetic circuits repress the *Caenorhabditis elegans* heterochronic gene *lin-28* after translation initiation. *Dev. Biol.* **243**, 215–225 (2002).
56. Kim, J. *et al.* Identification of many microRNAs that copurify with polyribosomes in mammalian neurons. *Proc. Natl Acad. Sci. USA* **101**, 360–365 (2004).
57. Doench, J. G., Petersen, C. P. & Sharp, P. A. siRNAs can function as miRNAs. *Genes Dev.* **17**, 438–442 (2003).
58. Saxena, S., Jonsson, Z. O. & Dutta, A. Small RNAs with imperfect match to endogenous mRNA repress translation: implications for off-target activity of siRNA in mammalian cells. *J. Biol. Chem.* **278**, 44312–44319 (2003).
59. Wassenaar, M. & Pelissier, T. A model for RNA-mediated gene silencing in higher plants. *Plant Mol. Biol.* **37**, 349–362 (1998).
60. Sijen, T. *et al.* On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* **107**, 465–476 (2001).
61. Schwarz, D. S., Hutvagner, G., Haley, B. & Zamore, P. D. Evidence that siRNAs function as guides, not primers, in the *Drosophila* and human RNAi pathways. *Mol. Cell* **10**, 537–548 (2002).
62. Roignant, J. Y. *et al.* Absence of transitive and systemic pathways allows cell-specific and isoform-specific RNAi in *Drosophila*. *RNA* **9**, 299–308 (2003).
63. Stein, P., Svoboda, P., Anger, M. & Schultz, R. M. RNAi: mammalian oocytes do it without RNA-dependent RNA polymerase. *RNA* **9**, 187–192 (2003).
64. Waterhouse, P. M., Wang, M. & Finnegan, E. J. Role of short RNAs in gene silencing. *Trends Plant Sci.* **6**, 297–301 (2001).
65. Gitlin, L. & Andino, R. Nucleic acid-based immune system: the antiviral potential of mammalian RNA silencing. *J. Virol.* **77**, 7159–7165 (2003).
66. Li, W. X. & Ding, S. W. Viral suppressors of RNA silencing. *Curr. Opin. Biotechnol.* **12**, 150–154 (2001).
67. Voynet, O. RNA silencing as a plant immune system against viruses. *Trends Genet.* **17**, 449–459 (2001).
68. Li, H., Li, W. X. & Ding, S. W. Induction and suppression of RNA silencing by an animal virus. *Science* **296**, 1319–1321 (2002).
69. Vargason, J. M., Szitty, G., Burgyn, J. & Hall, T. M. Size selective recognition of siRNA by an RNA silencing suppressor. *Cell* **115**, 799–811 (2003).
70. Ye, K., Malinin, L. & Patel, D. J. Recognition of small interfering RNA by a viral suppressor of RNA silencing. *Nature* **426**, 874–878 (2003).
71. Knight, S. W. & Bass, B. L. The role of RNA editing by ADARs in RNAi. *Mol. Cell* **10**, 809–817 (2002).
72. Tonkin, L. A. & Bass, B. L. Mutations in RNAi rescue aberrant chemotaxis of ADAR mutants. *Science* **302**, 1725 (2003).

73. Kennedy, S., Wang, D. & Ruvkun, G. A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. *Nature* **427**, 645–649 (2004).
74. Simmer, F. *et al.* Loss of the putative RNA-directed RNA polymerase RRF-3 makes *C. elegans* hypersensitive to RNAi. *Curr. Biol.* **12**, 1317–1319 (2002).
75. Calin, G. A. *et al.* Frequent deletions and down-regulation of microRNA genes *miR-15* and *miR-16* at 13q14 in chronic lymphocytic leukemia. *Proc. Natl Acad. Sci. USA* **99**, 15524–15529 (2002).
76. Michael, M. Z., O'Connor, S. M., van Holst Pellekaan, N. G., Young, G. P. & James, R. J. Reduced accumulation of specific microRNAs in colorectal neoplasia. *Mol. Cancer Res.* **1**, 882–891 (2003).
77. Moroy, T. *et al.* Structure and expression of *hcr*, a locus rearranged with c-myc in a woodchuck hepatocellular carcinoma. *Oncogene* **4**, 59–65 (1989).
78. Gauwerky, C. E., Huebner, K., Isobe, M., Nowell, P. C. & Croce, C. M. Activation of MYC in a masked t(8;17) translocation results in an aggressive B-cell leukemia. *Proc. Natl Acad. Sci. USA* **86**, 8867–8871 (1989).
79. Pfeffer, S. *et al.* Identification of virus-encoded microRNAs. *Science* **304**, 734–736 (2004).
80. Bohmert, K. *et al.* *AGO1* defines a novel locus of *Arabidopsis* controlling leaf development. *EMBO J.* **17**, 170–180 (1998).
81. Verdel, A. *et al.* RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672–676 (2004).
82. Zilberman, D., Cao, X. & Jacobsen, S. E. *ARGONAUTE4* control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299**, 716–719 (2003).
83. Pal-Bhadra, M., Bhadra, U. & Birchler, J. A. RNAi-related mechanism affects both transcriptional and posttranscriptional transgene silencing in *Drosophila*. *Mol. Cell* **9**, 315–327 (2002).
84. Kennerdell, J. R., Yamaguchi, S. & Carthew, R. W. RNAi is activated during *Drosophila* oocyte maturation in a manner dependent on *aubergine* and *spindle-E*. *Genes Dev.* **16**, 1884–1889 (2002).
85. Tijsterman, M., Okihara, K. L., Thijssen, K. & Plasterk, R. H. PPW-1, a PAZ/PIWI protein required for efficient germline RNAi, is defective in a natural isolate of *C. elegans*. *Curr. Biol.* **12**, 1535–1540 (2002).
86. Fagard, M., Boutet, S., Morel, J. B., Bellini, C. & Vaucheret, H. AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. *Proc. Natl Acad. Sci. USA* **97**, 11650–11654 (2000).
87. Dalmay, T., Horsefield, R., Braunstein, T. H. & Baulcombe, D. C. *SDE3* encodes an RNA helicase required for post-transcriptional gene silencing in *Arabidopsis*. *EMBO J.* **20**, 2069–2078 (2001).
88. Tijsterman, M., Ketting, R. F., Okihara, K. L. & Plasterk, R. H. RNA helicase MUT-14-dependent silencing triggered in *C. elegans* by short antisense RNAs. *Science* **295**, 694–697 (2002).
89. Domeier, M. E. *et al.* A link between RNA interference and nonsense-mediated decay in *Caenorhabditis elegans*. *Science* **289**, 1928–1931 (2000).
90. Mourrain, P. *et al.* *Arabidopsis* *SGS2* and *SGS3* genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell* **101**, 533–542 (2000).
91. Chan, S. W. *et al.* RNA silencing genes control *de novo* DNA methylation. *Science* **303**, 1336 (2004).
92. Smardon, A. *et al.* EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in *C. elegans*. *Curr. Biol.* **10**, 169–178 (2000).
93. Cogoni, C. & Macino, G. Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. *Nature* **399**, 166–169 (1999).
94. Glazov, E. *et al.* A gene encoding an RNase D exonuclease-like protein is required for post-transcriptional silencing in *Arabidopsis*. *Plant J.* **35**, 342–349 (2003).
95. Ketting, R. F., Haverkamp, T. H., van Luenen, H. G. & Plasterk, R. H. Mut-7 of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* **99**, 133–141 (1999).
96. Winston, W. M., Molodowitch, C. & Hunter, C. P. Systemic RNAi in *C. elegans* requires the putative transmembrane protein SID-1. *Science* **295**, 2456–2459 (2002).
97. Cogoni, C. & Macino, G. Posttranscriptional gene silencing in *Neurospora* by a RecQ DNA helicase. *Science* **286**, 2342–2344 (1999).
98. Kidner, C. A. & Martienssen, R. A. Spatially restricted microRNA directs leaf polarity through *ARGONAUTE1*. *Nature* **428**, 81–84 (2004).

Acknowledgements We thank W. Fischle, M. Poy and all members of the Tuschl laboratory for comments on the manuscript. G.M. is supported by an Emmy Noether fellowship of the Deutsche Forschungsgemeinschaft.

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

The functions of animal microRNAs

Victor Ambros

Dartmouth Medical School, Department of Genetics, Hanover, New Hampshire 03755, USA (e-mail: vra@dartmouth.edu)

MicroRNAs (miRNAs) are small RNAs that regulate the expression of complementary messenger RNAs. Hundreds of miRNA genes have been found in diverse animals, and many of these are phylogenetically conserved. With miRNA roles identified in developmental timing, cell death, cell proliferation, haematopoiesis and patterning of the nervous system, evidence is mounting that animal miRNAs are more numerous, and their regulatory impact more pervasive, than was previously suspected.

Noncoding' or 'non-messenger' RNAs are diverse molecules with structural, enzymatic and regulatory functions. Among those with regulatory activity are miRNAs, which are about 22 nucleotides in length and found in all metazoa studied so far¹. Of the 100–200 genes for distinct miRNAs¹ minimally contained in animal genomes, only a handful have known functions (Table 1). However, these functions suggest that miRNAs are important for the control of animal development and physiology.

As regulators of gene expression, miRNAs can work by essentially two modes^{2–7}. In plants, miRNAs base pair with messenger RNA targets by precise or nearly precise complementarity, and direct cleavage and destruction of the target mRNA through a mechanism involving the RNA interference (RNAi) machinery^{6,7} (see review in this issue by Meister and Tuschl, page 343). In contrast, most animal miRNAs are imprecisely complementary to their mRNA targets (Fig. 1), and they inhibit protein synthesis through an unknown mechanism that preserves the stability of the mRNA target: some studies even suggest that the translationally repressed target mRNAs remain associated with ribosomes^{8,9}.

Most miRNA genes seem to be solitary, and are expressed under the control of their own promoters and regulatory sequences. Other miRNA genes are arranged in clusters, and may be co-regulated with other members of the cluster. The *Drosophila* and *Caenorhabditis elegans* genomes each contain at least 100 different miRNA genes^{10–15}, and vertebrate genomes contain about 250 miRNA genes, as shown by complementary DNA cloning^{16,17} and computational predictions¹⁸. Some miRNAs are abundant, with an estimated 10,000 molecules per cell¹⁵, whereas others are only just detectable by hybridization to total RNA samples.

Here, we focus on recent findings from genomic experiments, whereby large numbers of miRNA genes are identified by cDNA sequencing and computational approaches; forward genetic experiments, whereby mutant genes are isolated from an organism showing abnormal phenotypic characteristics; and reverse genetic experiments, whereby a specific gene is knocked out (and/or overexpressed) to identify its function. These three approaches have identified roles for animal miRNA genes in the regulation of animal development and physiology (Fig. 2). Recently published computational predictions of potential targets of vertebrate and insect miRNAs will also be discussed.

Forward genetic analysis of miRNA function

Despite the success of cDNA cloning and bioinformatics approaches in identifying hundreds of miRNA genes, forward genetics remains one of the most fruitful approaches for

identifying miRNA genes with critical roles in the regulation of development and physiology (Fig. 2). The canonical miRNA genes, *lin-4* and *let-7* of *C. elegans*, were first identified by loss-of-function mutations that cause defects in developmental timing in the worm larvae^{19,20}. *lin-4* and *let-7* were then cloned on the basis of their mutant phenotypes, and the genes were found to encode small (21–22 nucleotide) noncoding RNAs that are processed from hairpin precursors. These hairpin precursors are a characteristic feature of the miRNA class of regulatory RNAs²¹. Regulatory targets of the *lin-4* and *let-7* miRNAs were identified from the genetic analysis of other genes with developmental timing phenotypes^{22–24}.

Two *Drosophila* miRNA genes have been identified by forward genetics. Screens for mutants defective in the regulation of programmed cell death and/or cell proliferation in the developing fly led to the identification of the *bantam* locus²⁵. When this was cloned, it was found to encode a miRNA²⁶. Similar genetic screens identified a locus affecting cell death and fat storage in the fly, which when cloned, was found to correspond to the *mir-14* gene²⁷. The *bantam* mutant phenotype included increased frequency of apoptosis. So genes with known roles in regulating the apoptotic pathway were prime candidates for *bantam* targets. The pro-apoptotic gene *hid* was discovered to contain sequences of partial complementarity to *bantam*, and *in vivo* tests supported a direct role for *bantam* in controlling Hid protein synthesis during development of the fly imaginal discs²⁶. The targets for *mir-14* in the control of apoptosis have not been identified yet, but may include cell-death effectors other than *hid*²⁷.

An exciting advance in the identification of developmental roles for animal miRNAs came from recent studies of genetic pathways controlling the asymmetric specification of certain neuronal cell types in *C. elegans*^{28–30}. Part of the worm's sensory discriminatory system, which allows worms to distinguish various attractive or repellent chemical stimuli in their environment, consists of two asymmetric chemosensory neurons: ASE left (ASEL) and ASE right (ASER). These two

Table 1 **Animal miRNA genes with genetically assigned functions**

miRNA	Animal	Function	Targets
<i>lin-4</i>	Ce	developmental timing ¹⁹	<i>lin-14</i> (refs 19, 22) <i>lin-28</i> (ref. 24)
<i>let-7</i>	Ce	developmental timing ²⁰	<i>lin-41</i> (ref. 23) <i>hbl-1</i> (refs 48, 49)
<i>lcy-6</i>	Ce	neuronal cell fate ²⁹	<i>cog-1</i> (ref. 29)
<i>mir-273</i>	Ce	neuronal cell fate ³⁰	<i>die-1</i> (ref. 30)
<i>bantam</i>	Dm	cell death, proliferation ²⁶	<i>hid</i> (ref. 26)
<i>mir-14</i>	Dm	cell death, fat storage ²⁷	caspase?
<i>miR-181</i>	Mm	haematopoietic cell fate ³³	?

Ce, *C. elegans*; Dm, *D. melanogaster*; Mm, *M. musculus*.

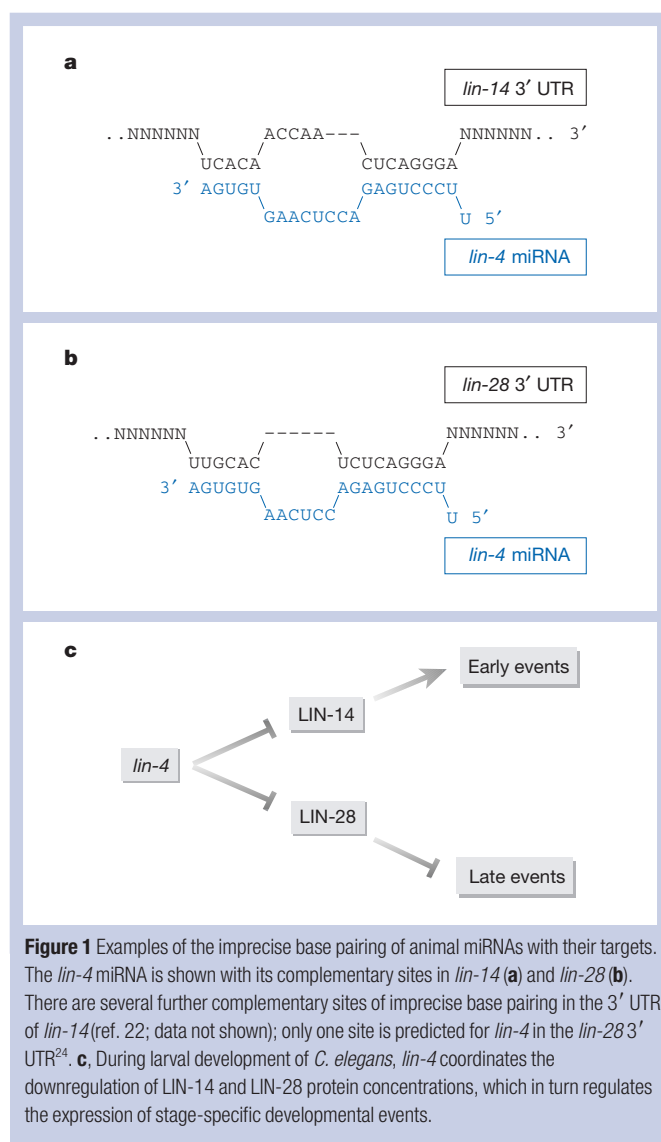


Figure 1 Examples of the imprecise base pairing of animal miRNAs with their targets. The *lin-4* miRNA is shown with its complementary sites in *lin-14* (a) and *lin-28* (b). There are several further complementary sites of imprecise base pairing in the 3' UTR of *lin-14* (ref. 22; data not shown); only one site is predicted for *lin-4* in the *lin-28* 3' UTR²⁴. c, During larval development of *C. elegans*, *lin-4* coordinates the downregulation of LIN-14 and LIN-28 protein concentrations, which in turn regulates the expression of stage-specific developmental events.

neurons detect different compounds, and together, allow the animal to respond selectively to stimuli. This behavioural specificity is in part rooted in differences in gene expression between ASEL and ASER. ASEL expresses one chemoreceptor gene, *gcy-7*, and ASER expresses another, *gcy-5* (Fig. 3). Genes encoding transcription factors (including *cog-1*, *lim-6* and *ceh-36*) that control the specification of ASEL and ASER asymmetry were identified by screens for mutants with either two ASEL cells or two ASER cells (as judged by *gcy-7::GFP* (where GFP is green fluorescent protein) or *gcy-5::GFP* expression²⁸). When one gene identified in these screens, *lgy-6*, was cloned, it was found to be unusual in that it does not encode a protein, but instead produces a novel miRNA²⁹ that regulates the synthesis of a downstream transcription factor COG-1. The predicted *lgy-6* transcript is a hairpin RNA with partial complementarity to a sequence in the 3' untranslated region (UTR) of *cog-1* mRNA: these sequences can mediate *lgy-6*-dependent repression of *cog-1* in transgenic worms²⁹.

The discovery of the *lgy-6* miRNA is remarkable in several respects. First, *lgy-6* probably could not be identified except by forward genetics. The *lgy-6* miRNA was not detected by cDNA cloning, and had not emerged from computational predictions of worm miRNAs^{11,31}, presumably because its relatively low contribution to worm total RNA meant its expression was not confirmed by northern-blot hybridization. Because *lgy-6* functions in a specific cell type, and a loss-of-function mutation results in a very subtle behavioural phenotype, the discovery of this particular miRNA gene required a

genetic screen that used ASEL- or ASER-specific GFP markers to directly assay for cell-fate transformations in mutants. Second, the difficulty of detecting such a scarce RNA as *lgy-6* by hybridization compelled Johnson and Hobert²⁹ to use unusual genetic experiments to prove that the *lgy-6* transcript forms a miRNA hairpin structure. In their approach, compensatory mutations were introduced into the *lgy-6* locus, and the mutant genes were then tested for function *in vivo*. The wild-type *lgy-6* sequence containing the putative miRNA hairpin structure rescued the *lgy-6* mutant phenotype, but a single point mutation designed to disrupt the predicted hairpin and prevent processing of the mature miRNA eliminated rescuing activity. Combining the original mutation with a compensatory mutation that was predicted to restore the hairpin structure restored rescuing activity. This provided strong genetic evidence that *lgy-6* functions as a hairpin RNA that is probably processed to a miRNA.

miRNAs found by genomics and reverse genetics

The discovery of *lgy-6* by forward genetics suggests that other interesting miRNA genes remain to be identified in *C. elegans* and other animals. An alternative to forward genetics for identifying rare miRNAs with developmental roles is to start with computationally-predicted miRNA genes and test for function by reverse genetics (Fig. 2). Remarkably, the latter approach has identified a second miRNA, *mir-273*, in the same pathway as *lgy-6* (ref. 30). *mir-273* is also a rare miRNA in the worm, and so was not found by cDNA cloning. Instead, *mir-273* was predicted computationally in a screen for phylogenetically conserved hairpin-forming sequences³¹. Computational methods for finding miRNAs are the ideal complement to cDNA cloning, which works best for relatively abundant miRNAs. However, a limitation of computational approaches is that predicted noncoding RNAs must be validated by an assay that can confirm expression of the RNA. Indeed, the number of worm miRNA candidates predicted computationally has far exceeded the number whose expression can be confirmed^{11,31}. A polymerase chain reaction with reverse transcription (RT-PCR)-based assay had to be used to detect the rare *mir-273* RNA in worm total RNA samples³¹.

The function of *mir-273* in worm neural development was discovered by the molecular characterization of the same pathway as that which involves *lgy-6* genes regulating ASER and ASEL sensory neuron asymmetry in the worm head. In addition to the *lgy-6* miRNA gene, and the *cog-1* homeobox target of *lgy-6*, several other genes were identified by mutations that alter the specification of ASEL and ASER. Among these was *die-1*, a gene encoding a zinc-finger transcription factor. The *die-1* promoter is active in both neurons, but far more DIE-1 protein (assayed using GFP-tagged DIE-1 expressed from a transgene) accumulates in ASEL than in ASER³⁰. Apparently, the *die-1* gene is regulated post-transcriptionally, and expressed specifically in ASEL. No candidate regulatory gene upstream of *die-1* was identified by mutation, but *die-1* mRNA was found to contain sequences that are complementary to parts of *mir-273*. Evidence that *mir-273* could be the repressor of *die-1* in ASER came from the use of transgenes expressing GFP from *mir-273* upstream sequences. In this way, *mir-273* regulatory elements that drove expression specifically in ASER were identified. No *mir-273* loss-of-function mutations were available, but ectopic expression of *mir-273* in ASEL, where its promoter is normally inactive, resulted in the transformation of ASEL into ASER³⁰.

The identification of two miRNAs in the ASEL/ASER pathway suggests that many developmental regulatory pathways, like the heterochronic and the ASEL/ASER pathways, involve several translational-regulatory steps involving miRNAs. Other miRNAs with very restricted expression patterns undoubtedly remain to be identified by forward genetics and carefully designed phenotypic assays, or by computational prediction followed by more sensitive expression assays.

cDNA cloning of rare, developmentally expressed miRNAs is aided by starting with RNA samples from material enriched for specific organs or cell types. For example, a set of six clustered mouse miRNAs

(*miR-290* to *miR-295*) are expressed specifically in embryonic stem (ES) cells³² and therefore, would not easily be cloned from differentiated tissues. Similarly, *miR-181* was identified by cloning from selected adult organs^{17,33}, and the role of *miR-181* in haematopoiesis development was indicated by its substantial enrichment in cDNA libraries made from mouse bone marrow and thymus³³. Consistent with its being cloned from these tissues, *miR-181* is expressed chiefly in bone-marrow B cells, and in thymus (presumably in T cells)³³.

Although *miR-181* loss-of-function mutations have not yet been isolated, Chen *et al.*³³ used a retroviral-based vector and an over-expression strategy to test whether *miR-181* specifically affects B-cell or T-cell development. Cells were extracted from mouse bone marrow, sorted by cell-type markers, and then infected with virus that had been engineered to overexpress *miR-181*, or other control miRNAs. Overexpression of *miR-181* in bone-marrow haematopoietic progenitor cells increased the numbers of B cells produced *in vitro* and *in vivo* and resulted in decreased numbers of CD8⁺ T cells³³. Although overexpression experiments must be interpreted with care when trying to infer the normal function of a gene, the Chen *et al.*³³ results suggest a role for *miR-181* in the normal development of B-cell and T-cell lineages in the mouse. *miR-181* may be a component of the trigger that directs bone-marrow progenitors along the B-cell differentiation pathway. For thymic T cells, *miR-181* seems to have a modulatory role in population homeostasis, perhaps as a sensor for signals that regulate the number of T cells. One prediction from these findings is that knockout of *miR-181* in haematopoietic lineages will lead to fewer B cells and an excess of thymic T cells.

miRNA gene clusters

A prominent characteristic of animal miRNAs is that their genes are often organized in tandem, and are closely clustered on the genome^{10,32}. In many cases, such clustered miRNAs are probably processed from the same polycistronic precursor transcript (a single mRNA molecule produced from the transcription of several tandemly arranged genes). When clustered miRNAs are of similar sequence, the cohort of gene products may contribute additively to the regulation of a set of mRNA targets. Clusters can also contain miRNAs of distinct sequences, suggesting that in these cases distinct miRNAs are coordinately deployed towards their various targets. For example, in insects and in vertebrate genomes, *let-7* and *miR-125* (orthologues of *C. elegans let-7* and *lin-4*, respectively) are closely clustered in tandem. *miR-125* and *let-7* are developmentally regulated together in flies^{34,35}, but in *C. elegans*, *lin-4* and *let-7* are unlinked, and are expressed sequentially^{36,20}. The clustered arrangement of *lin-4* and *let-7* in insects and vertebrates could represent an ancestral, conserved functional association between these two temporal regulators. A mammalian miRNA gene-cluster of particular interest encodes the six closely related genes for *miR-290* to *miR-295* in mouse³². These are expressed specifically in embryonic stem (ES) cells. Although it is not yet clear whether the human genome contains an orthologous cluster of ES-cell-enriched miRNAs, these findings suggest that translational repression of gene expression by miRNAs contributes to the maintenance of stem-cell potency³².

The properties of another cluster of miRNAs, on mouse distal chromosome 12 (and conserved at human 14q32), imply other interesting functions for miRNAs³⁷. This locus includes at least two miRNA genes (*miR-127* and *miR-136*) that are imprinted: they are expressed exclusively from the maternal chromosome. A retroviral transposon-like gene on the complementary strand to *miR-127* and *miR-136* is expressed only from the paternal chromosome. Consequently, *miR-127* and *miR-136* miRNAs are precisely complementary to the retroviral transposon-like mRNA. This arrangement suggests that *miR-127* and *miR-136* could regulate the transposon by means of an RNAi pathway³⁷. Although the function of this distal 12/14q32 cluster of miRNAs is unclear at present, the locus is of interest, if only because many other imprinted genes are known to have roles in

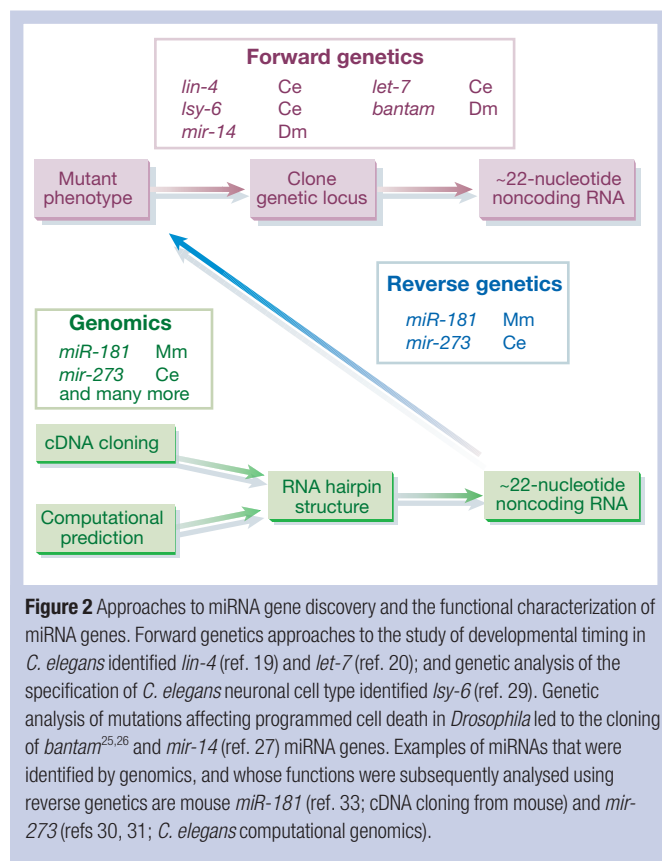


Figure 2 Approaches to miRNA gene discovery and the functional characterization of miRNA genes. Forward genetics approaches to the study of developmental timing in *C. elegans* identified *lin-4* (ref. 19) and *let-7* (ref. 20); and genetic analysis of the specification of *C. elegans* neuronal cell type identified *lzy-6* (ref. 29). Genetic analysis of mutations affecting programmed cell death in *Drosophila* led to the cloning of *bantam*^{25,26} and *mir-14* (ref. 27) miRNA genes. Examples of miRNAs that were identified by genomics, and whose functions were subsequently analysed using reverse genetics are mouse *miR-181* (ref. 33; cDNA cloning from mouse) and *mir-273* (refs 30, 31; *C. elegans* computational genomics).

early development. Moreover, a high concentration of other hairpin-forming miRNA candidates in the sequences surrounding *miR-127* and *miR-136* (ref. 37) suggests that many more mammalian miRNAs are yet to be discovered.

Prediction and validation of miRNA targets

The imprecise base pairing between the typical animal miRNA and a target mRNA suggests that any given miRNA can bind to a broad spectrum of different mRNAs, and so possess an enormous regulatory potential³⁸. This raises questions and concerns for systems biologists attempting to use computational approaches to identify miRNA target genes. Of all the predicted candidate targets of a given miRNA, which ones are authentic targets *in vivo*? The development of reliable methods for the computational prediction of animal miRNA targets necessarily involves an iterative process of algorithm design and refinement, guided by *in vivo* tests of the *in silico* predictions. So far, several published papers make good progress towards this end, although more work remains to be done.

Three groups have conducted computational screens for *Drosophila* miRNA targets on a genomic^{39,40} or subgenomic scale⁴¹, and two screens for vertebrate miRNA targets have been reported^{42,43}. All these approaches are based on the same fundamental assumption, which is that miRNA targets in general can be modelled on the base pairing between *lin-4* and *let-7* of *C. elegans* and their genetically-validated mRNA targets^{19,20,22–24}. The salient characteristics of this base pairing are: (1) the location of the miRNA complementary elements in 3' untranslated regions (UTRs) of mRNA targets, (2) the concentration of base pairing in a 'seed' or 'nucleus' of continuous Watson–Crick base pairing in the 5'–proximal half of the miRNA, and (3) the phylogenetic conservation of the complementary sequences in UTRs of orthologous genes. There is good evidence in support of the general applicability of these assumptions. First, for all the genetically-validated animal miRNA target genes, functional miRNA complementary sites lie within their 3' UTR sequences^{19,20,22–24,26,29,30}. Second, there are several lines of evidence,

apart from the topology of the known *C. elegans lin-4* and *let-7* sites, that point to the importance of base pairing in the 5' half of the miRNA¹. Of particular note are recent mutational studies using a variety of different miRNAs that support the primacy of critical base pairs and base-pairing patterns involving the first nine nucleotides of the miRNA^{43–45}. Finally, the criterion of phylogenetic conservation was found to be particularly important for reducing what is assumed to be the vast number of false-positive partial-complementary elements predicted for a given miRNA^{39–43}.

Despite the fundamentally similar approaches used for the published screens for miRNA targets, the results are not particularly congruent. This is probably because the approaches differ in certain important details, including: (1) the particular method of defining phylogenetic conservation, (2) the precise manner of modelling sites based on the verified cases, and (3) how thermodynamic and statistical factors are applied to score and rank predicted sites. As it is essentially impossible at this early stage to know which specific target-prediction method is more accurate, for now, it is probably wise to treat their results as complementary data sets.

Stark *et al.*³⁹ report the 100 highest-scoring targets for each of 73 known fly miRNAs, and Enright *et al.*⁴⁰ report the top 10 most statistically significant predicted targets for each fly miRNA. Of the 730 predicted targets reported by Enright *et al.*⁴⁰ about 15% were also flagged by Stark *et al.*³⁹ The predicted targets identified by both groups arguably represent the most reliable predictions from a biological standpoint. For example, both Enright *et al.*⁴⁰ and Stark *et al.*³⁹ identified genes of the Notch cell-surface receptor pathway as targets of *miR-7*. A third *Drosophila* miRNA target search did not survey all fly genes for miRNA targets, but focused on a selected set of protein genes that encode well-characterized regulators of embryonic patterning⁴¹. From among these patterning genes, 39 predicted targets of miRNAs were identified. About 20% of these targets were associated with the same miRNAs as those identified by Stark *et al.*³⁹, but none were among the top 10 lists from Enright *et al.*⁴⁰.

Lewis *et al.*⁴² identified orthologous genes from human, mouse and rat, which show conserved predicted sites for the same miRNA. A subset of these genes were identified as predicted targets for the same miRNA in mammals, and in the fish *Fugu rubripes*⁴². Kiriakidou *et al.*⁴³ reported several hundred predicted targets for mammalian miRNAs, and among these, two (initiation factor IF-2 targeted by *miR-20* and the protein ATAXIN targeted by *mir-101*) were also predicted by Lewis *et al.*⁴² One possible reason for the lack of overlap between the two vertebrate target screens is the difference in how the two groups dealt with target genes that are predicted to have just a single site for a given miRNA. Lewis *et al.*⁴² showed that the occurrence of multiple conserved sites for a given miRNA in the same UTR was a strong predictor of that gene being a statistically significant target. In contrast, genes with a single predicted high-scoring site could not, with confidence, be counted as potential miRNA targets. This was also the case in the insect target searches^{39–41}. However, Kiriakidou *et al.* made progress towards reliable prediction of single-site targets⁴³. Structure–function studies of a model miRNA complementary site and its cognate miRNA were conducted to derive rules for the characteristics of functional miRNA–target sites⁴³. These rules were incorporated into a computational algorithm, and the resulting rates of false-positive predictions (as judged by comparison to a population of randomized miRNA sequences) were low enough to permit statistically significant single-site predictions⁴³.

Increasing the statistical significance of single-site predictions is important, as there is evidence that multiple sites for the same miRNA are not necessarily required for targeting. For example, there is a single *lin-4* complementary site in the *C. elegans lin-28* 3' UTR²⁴. In fact, Kiriakidou *et al.* used the phylogenetically conserved *let-7b* site in the 3' UTR of human *lin-28* (ref. 46) as a test case for structure–function analysis, and showed that the site can mediate translational repression in human and mouse cultured cells⁴³. These sites seem to be absent from the insect *lin-28* (ref. 46), consistent with

evolutionary flexibility in miRNA–target–gene relationships. *let-7* sites in mammalian *lin-28* were not identified by Lewis *et al.*⁴², probably because the mouse and human *lin-28* 3' UTR contains too few repeats of these sites.

In vivo tests of some of the insect and vertebrate miRNA target predictions have produced promising results that tend to support the predictions. By fusing the UTR from predicted targets to a reporter construct, the regulatory capacity of the UTR-containing miRNA complementary elements could be assessed in transgenic flies^{39–40} or in animal cells^{42,43}. A significant fraction of the predicted UTR targets tested by these 'sensor' assays seemed to confer significant repression on the reporter, compared with a control UTR, in approximate agreement with the estimated frequencies of false positives^{39,40,42,43}. A problem with using these reporter tests to validate sites was that it was not always possible to control for the specificity of the miRNA, which would ideally involve comparing the reporter expression levels in the presence and absence of the miRNA.

It is worth noting that even when a predicted site seems to be validated by reporter studies, that site may not normally function *in vivo* (that is, if the miRNA and the mRNA are not normally co-expressed in the same cell). So, to derive a reliable picture of the likely miRNA pathways, it will be necessary to correlate the expression profiles of miRNAs with those of their potential target genes.

How might the current methods for computational prediction of miRNA targets be improved? Further *in vivo* analysis of predicted miRNA–target structures should lead to more refined rules of functional engagement, such as those derived by Kiriakidou *et al.*⁴³. If the short interfering RNA (siRNA)–transfection method^{43,45} is to become a method of choice for these studies, it is important to ensure that a miRNA introduced exogenously faithfully engages the normal endogenous miRNA-mediated translational repression^{8,9}. So far, the target prediction algorithms do not take into account the potential structure of the UTR sequences at or surrounding potential complementary sites. Using transgenic worms, Vella *et al.*⁴⁴ tested modified versions of the *C. elegans lin-41* 3' UTR, which is targeted by *let-7*, and found that sequences surrounding the miRNA sites are important for efficient *let-7*-mediated repression of *lin-41* translation. This finding reinforces the potential importance of incorporating into target prediction algorithms parameters that capture salient structural characteristics of the 3' UTR.

In future searches for animal miRNA targets, 5' UTR sequences should be included along with 3' UTRs, if only to critically test current assumptions about the primacy of 3' UTRs. Likewise, because plant miRNAs can recognize complementary elements within a coding sequence, future animal target searches should probably address this possibility directly. As discussed above, the assumption that miRNA–target binding is dominated by contiguous Watson–Crick base-pairing involving the 5'-proximal miRNA nucleotides seems to be reasonably well justified on theoretical and experimental grounds. However, there are likely to be important exceptions. For example, a site with nearly perfect complementarity to *miR-196a* is conserved in the 3' UTR of the *HOXB8* gene of mammals, fish and frog⁴⁷. But the site contains a conserved G·U base pair that breaks the 5' Watson–Crick 'seed' helix. Remarkably, the G·U base pair is the only imperfection in the *miR-196a::HOXB8* match, and it seems that this essentially continuous helix triggers cleavage of *HOXB8* mRNA *in vivo*, instead of translational repression⁴⁷. This result shows that in animals, as in plants, a miRNA–target match of near-perfect complementarity, even one containing G·U base-pairs in the miRNA 5' proximal block, can cause mRNA destruction.

The evolution of computational algorithms for animal miRNA target prediction will eventually lead to a scheme that is supported by experimental validation, that produces low rates of false-positive and false-negative predictions, and that reliably identifies functional single sites. Accurate rules for functional miRNA–target structures should permit the use of methods that are less dependent on conservation of the miRNA across wide phylogenetic distances. This is

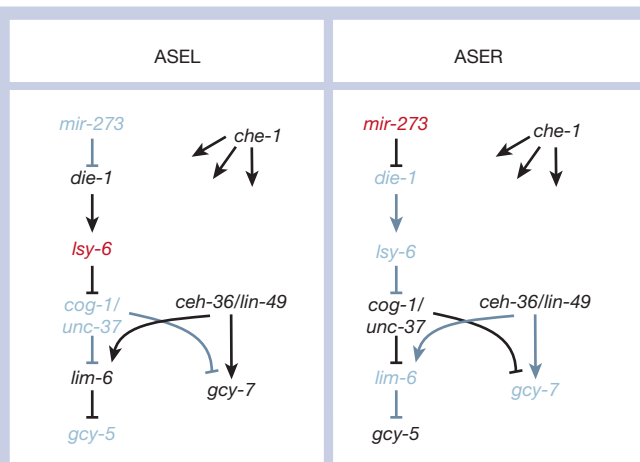


Figure 3 The roles of miRNAs *lsy-6* and *mir-273* in the pathway specifying the sensory neuron cell fates ASEL and ASER^{28–30}. The ASE cell type is specified in both ASEL and ASER by the expression of *che-1*, *unc-37* and *ceh-36/lin-49*. Distinct chemosensory functionalities of ASEL and ASER are defined by the transcription of the *gcy-7* or *gcy-5* chemoreceptors respectively. The asymmetric expression of *gcy-7* and *gcy-5* is specified by differential expression of upstream transcription factors, including *die-1*, *cog-1* and *lim-6*. *die-1* is translationally repressed in ASER by the *mir-273* miRNA³⁰, and *cog-1* is translationally repressed in ASEL by *lsy-6* miRNA²⁹. (Figure from Chang *et al.*³⁰.)

important for miRNAs such as those of the mouse ES-cell cluster, for which well-conserved human orthologues are not apparent. Some important miRNAs and/or miRNA targets may be less well conserved phylogenetically, and could be best identified by comparison of more closely related genomes, such as human and other primates.

The biological repertoire of miRNAs

What can be gleaned from the first set of miRNA/target searches regarding miRNA biology? Very few of the predicted vertebrate miRNA target genes were predicted to be targets for the orthologous miRNAs in insects. This suggests that although the sequence of a miRNA (and hence the sequence of its complementary sites) can be conserved across wide phylogenetic distances, the particular target genes (in which the binding sites reside) may be evolutionarily flexible. If binding sites (or blocks of UTR sequences containing the sites) can transpose between genes during evolution, then the biological pathways controlled by particular miRNAs would be evolutionarily plastic.

Are there any apparent trends in the types of gene that seem to be regulated by miRNAs? In *C. elegans*, the heterochronic genes targeted by *lin-4* or *let-7* encode nuclear proteins (LIN-14, HBL-1)^{19,48,49} and apparent RNA binding proteins (LIN-28, LIN-41)^{23,24,44}. *lsy-6* and *mir-273* regulate transcription factors^{29,30}. Similarly, the lists of predicted insect miRNA targets seem to be enriched in genes encoding transcription factors^{39–41}, but also include genes with diverse functions that are not directly related to gene expression. For example, the predicted targets of *miR-277* in *Drosophila* include a striking enrichment of genes in the biochemical pathway for the catabolism of leucine, isoleucine and valine⁴⁴. This result strongly suggests that *miR-277* could regulate this biochemical pathway at several points.

The control of cell fate is clearly a common theme for the activity of miRNAs¹. The heterochronic regulators *lin-4* and *let-7*, and the cell-death regulator *bantam*, control cell-fate choices for diverse cell types — ‘early’ versus ‘late’ in the case of the *lin-4/let-7* pathway, and proliferation versus death for the *bantam* pathway. *miR-181*, *lsy-6* and *mir-273* control more specific cell-fate choices for specific cell types. Similarly, *miR-181* may have distinct activities in the B-cell and T-cell lineages of mouse haematopoiesis, although the targets in these cells are not obvious from the published target searches.

Outlook

The genetic analysis of miRNA genes in model organisms is beginning to put into place the pieces of a mosaic that will eventually show us the range of functions that miRNAs have in the control of animal development and physiology. The computational prediction of miRNA–target interactions must work in parallel with genetics to identify miRNA-regulated pathways. These computational predictions will probably steadily improve in accuracy, as the existing algorithms are refined through an iterative process of *in silico* prediction and *in vivo* experimentation. Many outstanding questions about miRNAs remain. What affects the accessibility and efficacy of a miRNA at a UTR? What dictates whether an animal miRNA represses translation or directs mRNA cleavage? What is the nature and significance of functional cooperation and redundancy among miRNA genes? To what extent do distinct miRNAs act in combination on the same targets? The analysis of functional redundancy or epistatic relationships among miRNAs should be aided by the development of both genetic and nongenetic methods for the inhibition of particular miRNAs (singly or in combination). For example, recent publication of methods for the inhibition of miRNAs *in vivo* using complementary oligonucleotides^{50,51} should encourage the application of gene ‘knock-down’ approaches for the efficient analysis of miRNA function in cultured cells or intact animals.

The genetic analysis of miRNA function is an exciting challenge: the ‘miRNA milieu’ in a metazoan cell is likely to hold enormous potential for subtle and complex genetic regulatory interactions involving dozens of miRNAs and their numerous targets³⁸. Classical forward genetics continues to demonstrate its power in the miRNA arena, but much of what miRNAs accomplish may be both complex and subtle, demanding the creative application of genomics and reverse genetic approaches. □

doi:10.1038/nature02871

- Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
- Hutvagner, G. & Zamore, P. D. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**, 2056–2060 (2002).
- Zeng, Y., Wagner, E. J. & Cullen, B. R. Both natural and designed microRNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol. Cell* **9**, 1327–1333 (2002).
- Zeng, Y., Yi, R. & Cullen, B. R. MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc. Natl Acad. Sci. USA* **100**, 9779–9784 (2003).
- Doench, J. G., Peterson, C. P. & Sharp, P. A. siRNAs can function as miRNAs. *Genes Dev.* **17**, 438–442 (2003).
- Rhoades, M. W. *et al.* Prediction of plant microRNA targets. *Cell* **110**, 513–520 (2002).
- Tang, G., Reinhart, B. J., Bartel, D. P. & Zamore, P. D. A biochemical framework for RNA silencing in plants. *Genes Dev.* **17**, 49–63 (2003).
- Olsen, P. H. & Ambros, V. The *lin-4* regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev. Biol.* **216**, 671–680 (1999).
- Seggerson, K., Tang, L. & Moss, E. G. Two genetic circuits repress the *Caenorhabditis elegans* heterochronic gene *lin-28* after translation initiation. *Dev. Biol.* **243**, 215–225 (2002).
- Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**, 858–862 (2001).
- Lee, R. C. & Ambros, V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**, 862–864 (2001).
- Lagos-Quintana, M., Rauhut, R., Lendeckel, W. & Tuschl, T. Identification of novel genes coding for small expressed RNAs. *Science* **294**, 853–858 (2001).
- Ambros, V., Lee, R. C., Lavanway, A., Williams, P. T. & Jewell, D. MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Curr. Biol.* **13**, 807–818 (2003).
- Lai, E. C., Tomancak, P., Williams, R. W. & Rubin, G. M. Computational identification of *Drosophila* microRNA genes. *Genome Biol.* **4**(R42), 1–20 (2003).
- Lim, L. P. *et al.* The microRNAs of *Caenorhabditis elegans*. *Genes Dev.* **17**, 991–1008 (2003).
- Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W. & Tuschl, T. Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* **12**, 735–739 (2002).
- Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A. & Tuschl, T. New microRNAs from mouse and human. *RNA* **9**, 175–179 (2003).
- Lim, L. P., Glasner, M. E., Yekta, S., Burge, C. B. & Bartel, D. P. Vertebrate microRNA genes. *Science* **299**, 1540 (2003).
- Lee, R. C., Feinbaum, R. L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843–854 (1993).
- Reinhart, B. J. *et al.* The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**, 901–906 (2000).
- Ambros, V. *et al.* A uniform system for microRNA annotation. *RNA* **9**, 277–279 (2003).
- Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* **75**, 855–862 (1993).

23. Slack, F. J. *et al.* The *lin-41* RBCC gene acts in the *C. elegans* heterochronic pathway between the *let-7* regulatory RNA and the LIN-29 transcription factor. *Mol. Cell* **5**, 659–669 (2000).
24. Moss, E. G., Lee, R. C. & Ambros, V. The cold shock domain protein LIN-28 controls developmental timing in *C. elegans* and is regulated by the *lin-4* RNA. *Cell* **88**, 637–646 (1997).
25. Hipfner, D. R., Weigmann, K. & Cohen, S. M. The *bantam* gene regulates *Drosophila* growth. *Genetics* **161**, 1527–1537 (2002).
26. Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B. & Cohen, S. M. *bantam* encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell* **113**, 25–36 (2003).
27. Xu, P., Vernooij, S. Y., Guo, M. & Hay, B. A. The *Drosophila* microRNA *mir-14* suppresses cell death and is required for normal fat metabolism. *Curr. Biol.* **13**, 790–795 (2003).
28. Chang, S., Johnston, R. J. Jr & Hobert, O. A transcriptional regulatory cascade that controls left/right asymmetry in chemosensory neurons of *C. elegans*. *Genes Dev.* **17**, 2123–2137 (2003).
29. Johnston, R. J. & Hobert, O. A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature* **426**, 845–849 (2003).
30. Chang, S., Johnston, R. J. Jr, Frøkjær-Jensen, C., Lockery, S. & Hobert, O. MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode. *Nature* **430**, 785–798 (2004).
31. Grad, Y. *et al.* Computational and experimental identification of *C. elegans* microRNAs. *Mol. Cell* **11**, 1253–1263 (2003).
32. Houbaviy, H. B., Murray, M. F. & Sharp, P. A. Embryonic stem-cell-specific microRNAs. *Dev. Cell* **5**, 351–358 (2003).
33. Chen, C. Z., Li, L., Lodish, H. F. & Bartel, D. P. MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303**, 83–86 (2004).
34. Sempere, L. F., Sokol, N. S., Dubrovsky, E. B., Berger, E. M. & Ambros, V. Temporal regulation of microRNA expression in *Drosophila melanogaster* mediated by hormonal signals and broad-complex gene activity. *Dev. Biol.* **259**, 9–18 (2003).
35. Bashirullah, A. *et al.* Coordinate regulation of small temporal RNAs at the onset of *Drosophila* metamorphosis. *Dev. Biol.* **259**, 1–8 (2003).
36. Feinbaum, R. & Ambros, V. The timing of *lin-4* RNA accumulation controls the timing of postembryonic developmental events in *Caenorhabditis elegans*. *Dev. Biol.* **210**, 87–95 (1999).
37. Seitz, H. *et al.* Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nature Genet.* **34**, 261–262 (2003).
38. Bartel, D. P. & Chen, C.-Z. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nature Rev. Genet.* **5**, 396–401 (2004).
39. Stark, A., Brennecke, J., Russell, R. B. & Cohen, S. M. Identification of *Drosophila* microRNA targets. *PLoS Biol.* **1**, E60 (2003).
40. Enright, A. J. *et al.* MicroRNA targets in *Drosophila*. *Genome Biol.* **5**, R1 (2003).
41. Rajewsky, N. & Sock, N. D. Computational identification of microRNA targets. *Dev. Biol.* **267**, 529–535 (2004).
42. Lewis, B. P., Shih, L., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115**, 787–798 (2003).
43. Kiriakidou, M. *et al.* A combined computational-experimental approach predicts human microRNA targets. *Genes Dev.* **18**, 1165–1178 (2004).
44. Vella, M. C., Choi, E. Y., Lin, S. Y., Reinert, K. & Slack, F. J. The *C. elegans* microRNA *let-7* binds to imperfect *let-7* complementary sites from the *lin-41* 3' UTR. *Genes Dev.* **18**, 32–37 (2004).
45. Doench, J. G. & Sharp, P. A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504–511 (2004).
46. Moss, E. G. & Tang, L. Conservation of the heterochronic regulator Lin-28, its developmental expression and microRNA complementary sites. *Dev. Biol.* **258**, 432–442 (2003).
47. Yekta, S., Shih, L.-h. & Bartel, D. P. MicroRNA-directed cleavage of *HOXB8* mRNA. *Science*, **304**, 594–596 (2004).
48. Abrahante, J. E. *et al.* The *Caenorhabditis elegans* hunchback-like gene *lin-57/hbl-1* controls developmental time and is regulated by microRNAs. *Dev. Cell* **4**, 625–637 (2003).
49. Lin, S. Y. *et al.* The *C. elegans* hunchback homolog, *hbl-1*, controls temporal patterning and is a probable microRNA target. *Dev. Cell* **4**, 639–650 (2003).
50. Hutvagner, G., Simard, M. J., Mello, C. C. & Zamore, P. D. Sequence-specific inhibition of small RNA function. *PLoS Biol.* **2**, E98 (2004).
51. Meister, G., Landthaler, M., Dorsett, Y. & Tuschl, T. Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing. *RNA* **10**, 544–550 (2004).

Acknowledgements We thank members of the Ambros laboratory and numerous other colleagues for stimulating discussions. V.A. is supported by NIH grants.

Competing interests statement The author declares that he has no competing financial interests.

RNA silencing in plants

David Baulcombe

The Sainsbury Laboratory, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK (e-mail: david.baulcombe@sainsbury-laboratory.ac.uk)

There are at least three RNA silencing pathways for silencing specific genes in plants. In these pathways, silencing signals can be amplified and transmitted between cells, and may even be self-regulated by feedback mechanisms. Diverse biological roles of these pathways have been established, including defence against viruses, regulation of gene expression and the condensation of chromatin into heterochromatin. We are now in a good position to investigate the full extent of this functional diversity in genetic and epigenetic mechanisms of genome control.

Although RNA silencing has only emerged as a topic of general interest in the past six years, the first RNA silencing paper may have been published as long ago as 1928. In that paper Wingard described tobacco plants in which only the initially infected leaves were necrotic and diseased owing to tobacco ringspot virus¹ (Fig. 1). The upper leaves had somehow become immune to the virus and consequently were asymptomatic and resistant to secondary infection. At the time this 'recovery' was a mystery: there was no obvious way to explain the specificity of the resistance to secondary infection.

The details of the tobacco ringspot virus example remain to be worked out but we now know that recovery from virus disease involves RNA silencing that is targeted specifically at the viral RNA^{2,3}. There was no information about mechanisms in 1928 — it was not even known that the viral genome is RNA. But Wingard's paper is an appropriate starting point for the current interest in RNA silencing because it illustrates a viral defence role for RNA silencing which may have been one of its original functions in primitive eukaryotes. In modern plants this process has diversified into mechanisms that, in addition to defending the plant against viruses, protect the genome from transposons and regulate gene expression.

Here, I describe three natural pathways of RNA silencing in plants that have been revealed by genetic and molecular analysis. These pathways all involve the cleavage of a double-stranded RNA (dsRNA) into short 21–26-nucleotide RNAs by an enzyme Dicer that has RNase III domains. These RNAs are known as short interfering RNAs (siRNAs) and microRNAs (miRNAs). I discuss the possibility that there may be more than these three pathways, or that several variant mechanisms of RNA silencing exist. I also discuss features that distinguish plant miRNAs from animal miRNAs, and the amplification and mobile signal mechanisms in the siRNA pathways. Finally, I speculate about the role of RNA silencing in the integration of genome regulation.

Diverse RNA silencing pathways

The first pathway of the three is cytoplasmic siRNA silencing⁴. This pathway may be important in virus-infected plant cells where the dsRNA could be a replication intermediate or a secondary-structure feature of single-stranded viral RNA. In the case of plant DNA viruses, the dsRNA may be formed by annealing of overlapping complementary transcripts. The originally described recovery from tobacco ringspot virus¹ and many examples of transgene silencing in plants and animals are probably manifestations of cytoplasmic RNA silencing.

The second pathway is the silencing of endogenous messenger RNAs by miRNAs. These miRNAs negatively regulate

JOURNAL OF AGRICULTURAL RESEARCH

VOL. 37 WASHINGTON, D. C., AUGUST 1, 1928 No. 3

HOSTS AND SYMPTOMS OF RING SPOT, A VIRUS DISEASE OF PLANTS¹

By S. A. WINGARD²

Associate Plant Pathologist, Virginia Agricultural Experiment Station

INTRODUCTION



Figure 1 Recovery in tobacco plants infected with tobacco ringspot virus. The original legend¹ to the figure reads 'Turkish tobacco plant 23 days after inoculation with ringspot. Note the gradual decline in the development of ringspot symptoms on the upper leaves until finally the top leaves appear perfectly normal'. We now know that the virus causing the initial symptoms had activated viral RNA silencing that inhibited spread of the infection into the upper leaves, and caused them to be specifically immune to tobacco ringspot virus secondary infection.

gene expression by base pairing to specific mRNAs, resulting in either RNA cleavage or arrest of protein translation. Like siRNAs, the miRNAs are short 21–24-nucleotide RNAs derived by Dicer cleavage of a precursor. The prototype miRNAs in plants were identified as a subset of the short RNA population, with the molecular characteristics of the heterochronic RNAs *let-7* and *lin-4* in *Caenorhabditis elegans*⁵. miRNAs are derived from an inverted repeat precursor RNA with partially double-stranded regions, and they target a complementary single-stranded mRNA. However, there are differences between the miRNAs of plants and animals, as discussed in section ‘miRNA in plants’ below (see review in this issue by Ambros, page 350, for a more complete discussion of miRNAs in animals).

The third pathway of RNA silencing in plants is associated with DNA methylation and suppression of transcription. The first evidence for this type of silencing was the discovery in plants that transgene and viral RNAs guide DNA methylation^{6–8} to specific nucleotide sequences. More recently, these findings have been extended by the observations that siRNA-directed DNA methylation in plants is linked to histone modification⁹, and that, in fission yeast, hetero-chromatin formation at centromere boundaries is associated with siRNAs¹⁰. An important role of RNA silencing at the chromatin level is probably protecting the genome against damage caused by transposons (see review in this issue by Lippman and Martienssen, page 364).

An ancient origin of these three pathways of RNA silencing is likely because there are examples of each type in animals, fungi and plants. Green plants are unusual in that they have retained the capacity for all three types of silencing, whereas other organisms may have lost one or more of these pathways. Budding yeast, for example, has apparently lost all RNA silencing, and in mammals, all the examples of natural silencing found so far involve miRNAs⁵. Although exogenous RNAs in mammalian cells initiate ‘classical’ RNA interference (RNAi; a type of RNA silencing) involving siRNAs, it is not clear whether a specific siRNA pathway is involved. It could be that the exogenous RNAs are recruited into the miRNA pathway.

Argonaute and Dicer gene families

The Argonaute (Ago) proteins in plants, animals and fungi have been implicated in all three pathways of RNA silencing. In *Arabidopsis thaliana*, for example, *AGO1* mutants are defective for cytoplasmic RNA and miRNA silencing pathways, and *AGO4* mutants are impaired in chromatin silencing⁹. A central role of these AGO proteins seems likely because they are components of the silencing effector complexes that bind to siRNAs and miRNAs. Thus, the *Drosophila melanogaster* AGO2 protein binds siRNA by means of the PAZ (for piwi–argonaute–zwillie) domain¹¹, and is in the ribonuclease complex RISC (RNA-induced silencing complex)¹² that cleaves the target mRNA. Its role in RISC, on the basis of evidence with mouse AGO2 protein, is probably the ‘slicer’ ribonuclease in RISC¹³. AGO1 is probably a RISC component in *A. thaliana* because hypomorphic mutants retain the ability to accumulate miRNA, but the corresponding target mRNAs are not cleaved¹⁴. In fission yeast, an Ago protein is found in the RNA-induced transcriptional silencing complex (RITS) that targets heterochromatinization¹⁵. Finally, in *Tetrahymena*, an Ago homologue and siRNAs are found in a complex implicated in a silencing-related mechanism that leads to genome rearrangement^{16,17}.

This role in silencing effector complexes indicates that all silencing mechanisms will involve an Ago protein. Conversely it is likely that many if not all Ago proteins will be silencing-related. If this is the case, at least some of the ten Ago homologues in the *A. thaliana* genome may be associated with effector complexes of RNA silencing. Perhaps they associate with RISC or RITS that are adapted to silence genes in specialized cells, or at particular developmental stages. The mutant phenotype for *AGO7* in *A. thaliana*, for example, is altered timing of the phase change between juvenile and adult leaves: so AGO7 might be part of an RNA silencing effector complex which is specific to a particular stage of development¹⁸.

The diverse RNA silencing pathways may not be completely separate. Some proteins, for example AGO1 and zwillie (AGO10), have partially overlapping functions. Conversely, as illustrated by the miRNA- and cytoplasmic-siRNA-defective phenotype of *ago1* mutants in *A. thaliana*^{14,19}, a single Ago protein may participate in multiple silencing pathways. Mutations at another silencing-related gene, *HEN1*, also indicate overlap in different silencing pathways: *HEN1* mutants are defective in both miRNA and cytoplasmic siRNA silencing²⁰.

The Dicer gene family in *A. thaliana* has only four members²¹: presumably, if there are more than four silencing pathways involving Ago proteins, some Dicers will be active in more than one of them. Two Dicer proteins in *A. thaliana* have well defined functions: Dicer-like (DCL)1 is required for miRNA biogenesis^{22–24}; DCL3 produces retroelement and transposon siRNAs and is required for chromatin silencing²⁴. The DCL3 products correspond to a class of siRNA that, from earlier analyses, had been associated with transposons²⁵, and that are longer (24 nucleotides rather than 21 nucleotides)^{25,26} than the typical DCL1 products. However the role of the other two Dicers, DCL2 and DCL4, has been more difficult to define. DCL2 has been implicated in viral siRNA production but the loss-of-function phenotype is only a transient reduction in the level of siRNA in one of several viruses tested²⁴. It is likely, therefore, that there is functional redundancy and that the other Dicers of *A. thaliana* are also involved in viral siRNA production. The function of DCL4 is not known.

miRNAs in plants

Plant and metazoan miRNA pathways are fundamentally the same: both involve 20–22-nucleotide single-stranded miRNAs that are generated by a Dicer and both depend on an Ago protein. However, the metazoan miRNAs are processed by Drosha and Dicer RNase III in two steps that take place in the nucleus and cytoplasm⁵, whereas miRNAs in plants are processed by a Dicer, and this is most likely to occur in the nucleus²². Associated with this processing difference, a dsRNA binding protein, HYL1, is specific to the plant miRNA pathway^{27,28}. A further important difference is that the plant miRNAs are more perfectly paired to their target RNA and use RNA cleavage rather than translation suppression as the primary silencing mechanism^{29–31}. The animal miRNAs are normally targeted to the 3′ untranslated region (UTR) of a mRNA, whereas the plant miRNAs have targets in the coding sequence or even in the 5′ UTR³².

There are now extensive lists of plant miRNAs (<http://cgrb.orst.edu/smallRNA/> and <http://www.sanger.ac.uk/Software/Rfam/mirna/>) and, in several cases, the target mRNA has been validated experimentally by expression of an miRNA-resistant target gene with silent mutations in the putative miRNA complementary region (Fig. 2). The mutations interfere with miRNA targeting and result in overexpression of a bona fide miRNA target. This gold standard of miRNA target identification has been applied to target mRNAs encoding: (1) a TCP transcription factor required for leaf morphogenesis³³; (2) a MYB33 transcription factor involved in a plant hormone response³³; (3) an HD-ZIP transcription factor that influences abaxial and adaxial polarity in leaves and stems³⁴; (4) the AP2 transcription factor that regulates floral development³⁵; (5) a NAC transcription factor that plays a role at many stages of development³⁶; and (6) the AGO1 cofactor of silencing¹⁴. Other examples that have been validated by the detection of target mRNA cleavage products corresponding to an miRNA-complementary site include a scarecrow transcription factor of unknown function³⁷ and DCL1 (ref. 38).

These, and many other of the miRNAs in *A. thaliana* for which the targets have been identified computationally, correspond to mRNAs for transcription factors and other proteins involved in developmental regulation^{30,31,39}. The mRNAs for proteins associated with ubiquitin-mediated protein degradation are also potential miRNA targets^{30,32}. The target sequences in most of these examples are conserved in rice and *A. thaliana* and, in one instance, the conservation extends even

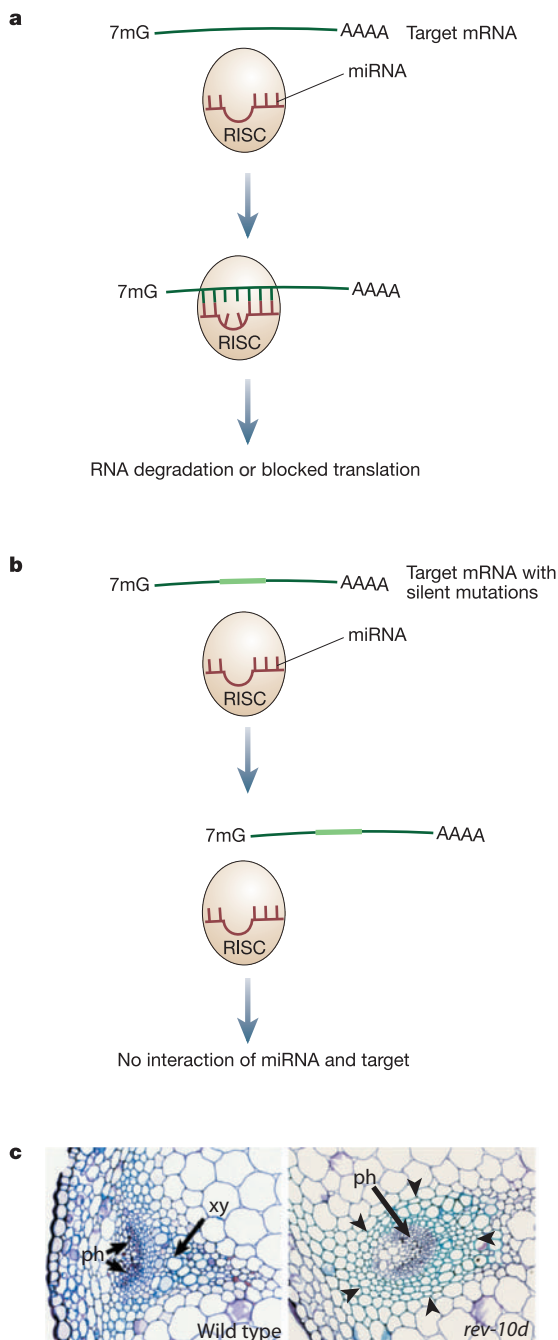


Figure 2 Validation of miRNA targets. In a wild-type plant (a) an miRNA associated with RISC will base pair to its cognate target and promote either sequence-specific RNA degradation or a translational block. However, if a transgene is introduced in which the miRNA target sequence has silent mutations (b), the miRNA cannot bind to the target sequence and the protein encoded by the mRNA is overexpressed. c, Cross-sections of stems in wild-type and *rev-10d* transgenic *A. thaliana*, illustrating a phenotype from a miRNA-resistant mRNA (from ref. 34). The *rev10d* transgene encodes the revoluta transcription factor and its RNA is resistant to targeting by *miR165* and *miR166*. In the wild-type plants the xylem (xy) is positioned centrally and inside the peripheral phloem (ph) tissue. In the *rev-10d* stems the vascular bundles are radialized with xylem tissue (arrowheads) surrounding phloem tissue (ph). This effect on the distribution of xylem tissue implies that interpretation of positional information requires correct targeting by *miR165* and *miR166*. 7mG, 7-methylguanine.

between mosses and flowering plants⁴⁰. However, the range of targets is not restricted to 'developmental' genes because there are also miRNAs that increase or decrease in abundance following cold or drought stress or sulphur starvation. The predicted targets of these miRNAs encode a more diverse set of proteins including laccases, cytochrome *c* oxidases, spliceosomal proteins and ATP sulphurylases^{30,32}.

It has been estimated that the *A. thaliana* genome has about 100 miRNA loci³⁰. This estimate, however, is based on a computational genome survey which assumes that miRNA targets are conserved in *A. thaliana* and rice. Putative miRNA targets that are conserved between *A. thaliana* and *Lotus*, *Medicago* or *Populus* and not rice³², or that are not conserved in distantly related species, would not have been identified in this survey and the number of miRNA loci could be considerably higher.

Initiation and amplification of silencing

RNA-dependent RNA polymerases (RDRs; also known as RdRPs) are required for the cytoplasmic and chromatin RNA silencing pathways in *C. elegans*^{41,42}, fungi^{10,43} and plants^{24,44,45} but not, apparently, for the same pathways in insects or mammals. The RDRs share a common sequence motif that is distantly related to the catalytic domain of DNA-dependent RNA polymerases⁴⁶, and it is therefore likely that they are an ancient group of proteins. *A. thaliana*, *C. elegans* and *Neurospora crassa* have small RDR gene families that, as with Ago proteins, indicate functional diversification of silencing pathways. In *A. thaliana* the RDR1 and RDR6 (also known as SDE1/SGS2) orthologues are required in the cytoplasmic RNA silencing pathway that silences transgenes and viruses. However, it seems that these proteins have specificity for different viral RNAs: RDR6 mutants in *A. thaliana* are hypersusceptible to cucumber mosaic virus⁴⁵ but not to tobacco rattle and tobacco mosaic virus⁴⁷, whereas tobacco plants with reduced levels of RDR1 show enhanced susceptibility to tobacco mosaic virus⁴⁸. RDR2 mutants are defective for production of endogenous siRNAs, including those corresponding to retroelements. So the RDR2 protein might be part of the chromatin silencing pathway²⁴.

In principle, the RDR proteins could mediate primer-dependent and primer-independent mechanisms of RNA silencing (Fig. 3). The primer-independent process may be important for the production of dsRNA from a single-stranded template, so that silencing can be initiated in virus-infected plants or with transgene RNAs (Fig. 3a). Consistent with this primer-independent mechanism, *in vitro* assays with *N. crassa*⁴⁹ and tomato enzymes⁵⁰ demonstrate that RDR catalyses primer-independent synthesis of dsRNA on a single-stranded RNA (ssRNA) template. Similarly, in wheat germ extracts, ssRNA can be copied into complementary RNA by an unidentified enzyme that, presumably, is an RDR²⁶. However, it is not yet clear how RDR could differentiate the viral and transgene RNAs targeted for silencing from the non-silenced endogenous RNAs. Perhaps the RNA that becomes silenced contains 'aberrant' features that are absent from 'normal' non-silenced RNA. Alternatively, the aberrant RNA might lack features that are present in normal RNA. The absence of a 5' cap (R. Sablowski, personal communication) renders an RNA susceptible to RDR-dependent RNA silencing in *A. thaliana*, but other possibilities have not been ruled out.

The second RDR mechanism (Fig. 3b) requires that primary siRNAs from a virus, transposon or transgene are primers in RDR-directed synthesis of dsRNA. The QDE1 RDR protein from *N. crassa* incorporates a labelled 20-nucleotide antisense RNA into the complementary strand of a ssRNA *in vitro*⁴⁹, in a manner that is consistent with this mechanism. This primer-dependent process is also supported by indirect genetic evidence from *C. elegans* and plant systems in which the initiator of silencing comes from part of a target gene. In these systems, the secondary siRNAs that accumulate in the silenced tissue are dependent on RDR proteins^{41,51}, and are derived not only from the initiator region but also from adjacent regions in the target sequence. Moreover, the secondary siRNAs in *C. elegans*

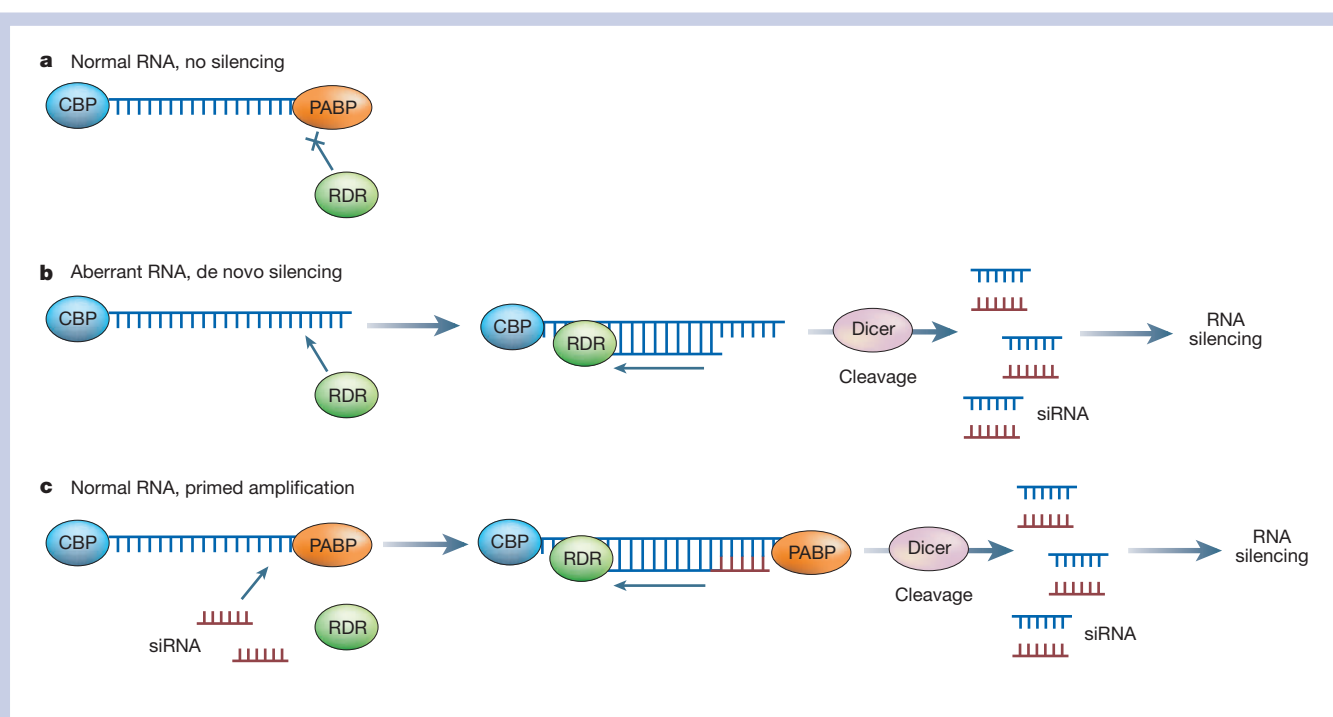


Figure 3 The action of RDR proteins in the initiation or amplification of silencing. **a**, RNAs are normally not silenced because the RDR proteins do not have access to the template RNA sequence. Cap-binding protein (CBP) and poly-adenosine-binding protein (PABP) may be involved in this restriction of RDR access. However in **b**, the RDR protein is allowed access because the RNA lacks a 5' cap or 3' poly-adenosine tail, and dsRNA is produced which enters the siRNA pathway. **b**, The amplification process would result from the ability of a single aberrant RNA to generate many molecules of siRNA. **c** shows the outcome if a small quantity of primary siRNA is present from either a virus, a transposon or from a cellular RNA through the process shown in **b**. The antisense strand of this siRNA may anneal by base pairing to a target RNA and serve as a primer for the RDR. The resulting dsRNA would then be cleaved by Dicer and, as in **b**, there would be amplification because many secondary siRNAs would be produced from each molecule of primary siRNA.

correspond only to the 5' side of the ssRNA, as would be expected if an antisense primary siRNA had been extended by the RDR at its 3' end. However, in *A. thaliana* and *Nicotiana benthamiana* the secondary siRNAs are from both the 5' and the 3' side of the initiator^{51,52} on the ssRNA, and so cannot be produced from a simple priming mechanism on a single RNA species. The most likely explanation here is that the silencing target, like many parts of the *A. thaliana* genome, is transcribed from both strands⁵³. The 3' secondary siRNAs would then result from extension of an siRNA primer on an antisense RNA template.

As a result of the RDR-mediated mechanisms, a single aberrant RNA species or primary siRNA molecule could generate many dsRNAs which would then silence even more target molecules. This amplification process is likely to be essential in virus defence because it would ensure that silencing of viral RNAs keeps pace with the replication and accumulation of viral RNA. Similarly, in genome defence, the amplification steps would ensure that a few molecules of transposon RNA could activate the chromatin-silencing pathway sufficiently to suppress all copies of a transposable element. In addition, the RDR proteins would help target the RNA silencing mechanism to transposons because transcripts with direct repeats are readily amplified⁵⁴ (see review in this issue by Lippman and Martienssen, page 364).

Mobile silencing signals

Together with RDR amplification, mobility of a silencing signal is probably a crucial characteristic of an antiviral defence system. A mobile silencing signal could move either with or ahead of the virus to silence the viral RNA before, or at the same time, as the virus moves into a cell⁵⁵. Indeed, in plants and *C. elegans*, the effects of silencing extend beyond those cells in which the silencing is initiated and can spread systemically^{56–58} through the organism. This systemic effect has nucleotide-sequence specificity corresponding to the initiator

dsRNA, indicating that the signal either is an RNA or that it has an RNA component. Consistent with this idea, the SID1 protein, which is required for systemic RNAi in *C. elegans*, is a transporter of dsRNA across membranes^{59,60}.

In plants the systemic silencing mechanism is unlikely to be the same as that in *C. elegans*. The signal does not have to cross any membranes because most of the cells in a plant, including the phloem cells of the vascular system, are connected by plasmodesmatal channels that are a continuation of the endoplasmic reticulum⁶¹. So far, none of the host proteins involved in movement of this silencing signal has been identified. However, an analysis of systemic signalling from a green fluorescent protein (GFP) transgene coupled to a phloem-specific promoter indicated that the signalling mechanism in plants can be resolved into short (up to 15 cells) and longer range⁶² phases extending up to several centimetres.

Short-range signalling is unlike the longer range movement because it is unaffected by *RDR6* loss-of-function mutants (Fig. 4), and it is likely that a 21-nucleotide siRNA is the mobile signal⁶². Consistent with a short RNA being the mobile signal for short-range signalling, the siRNA in a virus-infected cell is present either as free RNA or in low molecular weight complexes that could be well below the normal size exclusion limit of plasmodesmata⁶³.

A longer 24-nucleotide class of siRNAs, possibly generated by DCL3 (see section 'Argonaute and Dicer gene families' above), has been proposed as a candidate for the long-range phloem entry signal because viral proteins that block systemic silencing also prevent accumulation of the 24-nucleotide siRNA²⁵. However, systemic silencing is transmitted from grafted plants in which both the 21- and 24-nucleotide siRNAs are suppressed by the viral HCPro suppressor of silencing⁶⁴. It is therefore possible that other silencing RNAs including long ssRNA, dsRNA or siRNAs, could be signal molecules because any of them can initiate silencing if they are introduced into a cell with a suitable target. The plasmodesmatal size exclusion limit⁶¹

might be a barrier to intercellular movement of high molecular weight complexes containing these RNAs, but it is possible that free RNAs or RNAs in low molecular weight complexes are mobile. Precedents for phloem-mobile RNAs in plants include viroids⁶⁵ — small non-protein coding RNA plant pathogens — and the transcript corresponding to a homeobox protein mRNA⁶⁶.

An intriguing possibility is that the movement of miRNAs and endogenous siRNAs might play a role in regulation of endogenous genes. For example, during leaf development *miR165* and *miR166* are negative regulators of genes affecting leaf polarity, and their distribution and possible gradient of expression is consistent with that of a mobile signal^{34,67,68}. Consistent with the idea that endogenous silencing RNAs are mobile signals, there are many miRNAs and siRNAs in the phloem sap of pumpkin⁶⁹, and a phloem protein has been detected that binds specifically to the single-stranded form of these RNAs.

Viral suppressor proteins

RNA silencing in plants prevents virus accumulation and, accordingly, viruses have evolved various strategies to counteract this defence mechanism. The primary counter defence measure involves suppressor proteins of silencing which are encoded in the genomes of both RNA and DNA viruses⁷⁰. These proteins probably evolved independently in different virus groups because they are structurally diverse, and there are no common sequence motifs. A secondary mechanism to counteract silencing is illustrated by the apparent resistance of satellite and defective interfering RNAs to degradation by siRNAs^{71,72}. It seems that these RNAs have protective secondary structures, or are compartmentalized so that they are hidden from the RNA silencing mechanism.

In principle, the plant viral suppressor proteins could be used as a tool to investigate the mechanism of RNA silencing. However, in most instances, including the prototype HCPro suppressor from potyviruses, there are conflicting data about their mechanism of action^{73–75}. Only for two suppressors — p21 encoded by beet western yellow virus⁷³ and p19 encoded by the tomato bushy stunt virus (TBSV) group^{63,74} — is there a clear indication of how they act. In both cases, the virus suppressor protein binds to and, presumably, inactivates siRNAs so that they do not target the corresponding viral RNAs. For p19, the high resolution crystal structures for two different TBSV-group proteins, combined with molecular and biochemical data, indicate precisely how silencing is blocked. A tail-to-tail p19 homodimer forms α -helix brackets around the ends of the siRNA base-paired region^{76,77} and, consequently, an siRNA or miRNA is prevented from being incorporated into an active RISC^{63,74,78}. In transgenic *Arabidopsis* expressing p19 (ref. 73), both miRNA and its complement (miRNA*) accumulate, whereas in the control plants without p19 the miRNA* is undetectable⁷⁴. Presumably the miRNA–miRNA* duplex is normally a short-lived precursor of miRNA–RISC⁷⁹, but is stabilized in the presence of p19.

Given the likelihood that virus defence was an ancient role of RNA silencing, it would not be surprising if RNA silencing also influences animal virus infections. Consistent with this idea, the NS1 and E3L proteins of influenza and vaccinia viruses⁷⁸, and the B2 protein of flock house virus⁸⁰, have silencing suppressor activity. In addition, there are five different miRNAs in mammalian cells infected with Epstein–Barr virus that correspond to inverted repeat regions in the viral genome⁸¹. One of these viral miRNAs targets the viral DNA polymerase gene *BALF5*, and directs processing of a *BALF5* transcript which is likely to affect virus accumulation. However, this is the only report of siRNAs or miRNAs corresponding to mammalian viruses, and it remains to be confirmed whether the silencing suppressor activity of the influenza and vaccinia virus proteins is a side effect of their dsRNA binding activity⁸². So, on the basis of current evidence, it seems unlikely that RNA silencing in mammals is a general defence mechanism against viruses as it is in plants. Perhaps it is effective against a subset of mammalian viruses or is an antiviral defence in

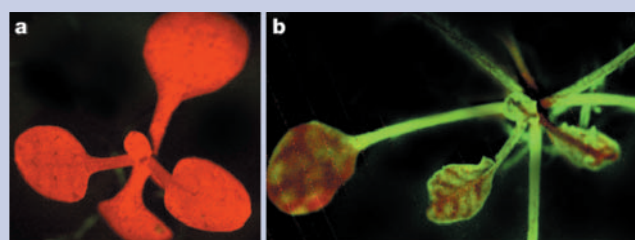


Figure 4 A silencing signal is affected by an RDR mutation. The two panels show *A. thaliana* plants carrying two transgenes. **a**, In the wild-type plant, a GFP transgene is constitutively transcribed, but the GFP fluorescence is suppressed because a second transgene for GFP dsRNA is expressed in the phloem cells. A silencing signal has moved out of the phloem and has silenced the GFP transgene throughout the leaf. The plant appears red under ultraviolet light owing to chlorophyll fluorescence. **b**, This plant has a mutation in *RDR6* and the silencer signal is able to act only in the cells that are close to the phloem. The GFP transgene is not silenced in cells that are further than about 20 cells from the phloem and consequently they appear green under ultraviolet light (reproduced with permission from ref. 62.).

embryonic or other cells in which the systems of innate and humoral immunity are ineffective.

Viral symptoms and silencing

As viruses are inducers, suppressors and targets of the RNA silencing mechanism, there are many ways in which the symptoms in infected plants can be influenced by viral intervention in the miRNA and siRNA pathways. For example, in *A. thaliana* plants infected with turnip mosaic virus, the symptoms include developmental defects that are mimicked by transgenic expression of the HCPro suppressor of silencing, or by mutation affecting the DCL1 protein of the miRNA pathway⁸³. It is likely, therefore, that these symptoms are caused by suppression of the host's miRNA pathway by the viral HCPro. By extrapolation, other viral symptoms involving developmental defects are probably due to silencing suppressors^{33,34,68}.

A variation on this effect of viral suppressors is suggested by the finding that transgenic tobacco plants expressing HCPro show enhanced resistance to diverse pathogens including tobacco mosaic virus and the oomycete (water mould) *Peronospora tabacina*. One plausible explanation for this resistance is that HCPro suppresses the action of endogenous miRNAs or siRNAs that usually target negative regulators of the host's innate immune system⁸⁴.

A second more direct role of RNA silencing in symptom formation is illustrated by cucumber mosaic virus strains with small non-coding satellite RNAs. The Y strain of satellite RNA causes a bright yellow chlorosis in infected plants⁷² that is suppressed in transgenic plants expressing the HCPro silencing suppressor. Presumably, the symptoms develop because there are Y-satellite-RNA-derived siRNAs that target an endogenous gene normally required for chlorophyll accumulation. Similarly in viroid infections, pathogen-derived siRNAs may target endogenous genes⁷². Viroid RNAs would be an excellent substrate for Dicer because they are essentially circular RNAs that fold to form an extensively double-stranded rod-like structure. Consistent with the involvement of siRNAs, viroid symptoms are blocked if the infected plant expresses HCPro, and are mimicked in transgenic plants expressing an inverted repeat transgene derived from the symptom-inducing region of the viroid genome⁷².

RNA silencing in genetic regulation

One of the challenges for research in genetic regulation is to understand how expression of whole sets of genes can be coordinated. RNA silencing may contribute to this integration of genetic control because it is subject to several levels of feedback regulation, and because any one siRNA or miRNA could target the silencing mechanism to multiple RNAs or to DNA loci. The combined effect of these features could be

that a single RNA species mediates RNA silencing-based effects on many other genes and RNAs. One level of feedback control is illustrated by cytoplasmic RNA silencing in which dsRNA is generated from a ssRNA by an RDR (Fig. 3a). In this scenario the original ssRNA template is both a target and a precursor of the siRNA. High levels of the ssRNA would therefore lead to abundant siRNA and, consequently, the ssRNA levels would decline. Conversely, reduced levels of the ssRNA would lead to decreased amounts of siRNA and ultimately an increase in the amount of ssRNA (Fig. 5a).

Feedback mechanisms are also apparent in the miRNA pathways because the mRNA transcripts encoding DCL1 (ref. 38) and the AGO1 component of RISC¹⁴ are themselves targets of miRNAs (*miR162* and *miR168*, respectively). Abundant AGO1 or DCL1 proteins would lead to a silencing-mediated decrease in the amount of corresponding mRNAs. Conversely, reduced amounts of these proteins would ease the level of silencing and the concentration of mRNAs would increase. This feedback mechanism could explain the otherwise paradoxical increase in the amount of miRNA in the presence of viral suppressors of silencing^{73–75} (Fig. 5b). In this case, the suppression of silencing would uncouple the feedback loop so that the abundance of AGO1, DCL1 and the associated miRNAs would be unchecked by the normal mechanisms.

A second type of feedback control is implied by the finding that *miR159* and its putative target (transcription factor *MYB33* mRNA) are both positively regulated by the plant hormone gibberellic acid (GA)⁸⁵. A GA stimulus could lead to an increase in *MYB33* that would initiate flowering and, directly or indirectly, to an increase in *miR159*. The higher concentration of *miR159* would then counteract the increase in *MYB33* and dampen the GA response (Fig. 5c). A similar mechanism may apply to *miR171* that targets the transcription factor *GRAS* mRNA³¹. If *miR171* were a simple negative regulator of the transcription factor mRNA then the miRNA would be abundant when the target is rare and vice versa. In fact both are upregulated in

inflorescences³¹ in a pattern that could be explained if the *GRAS* transcription factor promoted expression of the *miR171*.

The ability of siRNAs and miRNAs to target multiple RNAs and DNAs may also contribute to their role in integration of genetic regulation. Targeting by these short RNAs is tolerant of limited sequence mismatches⁵, and each siRNA or miRNA probably has both primary and secondary targets. The primary targets would be more completely matched to the siRNA or miRNA and, if the target is an RNA, it would be cleaved by RISC. Any secondary targets would have more mismatches and would be cleaved more slowly than the primary targets. Alternatively, as occurs for *APETALA2* (*AP2*) mRNA^{35,86}, the secondary RNA targets might be translationally repressed rather than degraded. Currently there is no information about secondary targets of siRNAs and miRNAs in plants. Investigation of *miR159* and *miR319* may tell us about the potential importance of secondary targets because the two miRNAs differ at only three nucleotide positions and have *MYB* or *TCP* mRNAs, respectively, as distinct primary targets³³. The extent to which these miRNAs cross target at either the RNA cleavage or translational repression level will indicate the extent to which miRNAs in plants might have multiple targets.

Future prospects

Over the past few years, we have come to appreciate that there are diverse natural roles of RNA silencing, ranging from defence against viruses to the regulation of gene expression and chromosome structure. But a remaining challenge is to find out the full extent of this functional diversity. One approach will be to analyse Ago and other gene families associated with the silencing mechanism. The characterization of silencing-related RNAs using computational approaches and direct sequence analysis may also be informative, as is discussed above in the context of miRNAs. In addition to miRNAs, plants including *A. thaliana* have vast numbers of other short RNA species, many of which are likely to be siRNAs^{23,28}. A small RNA database

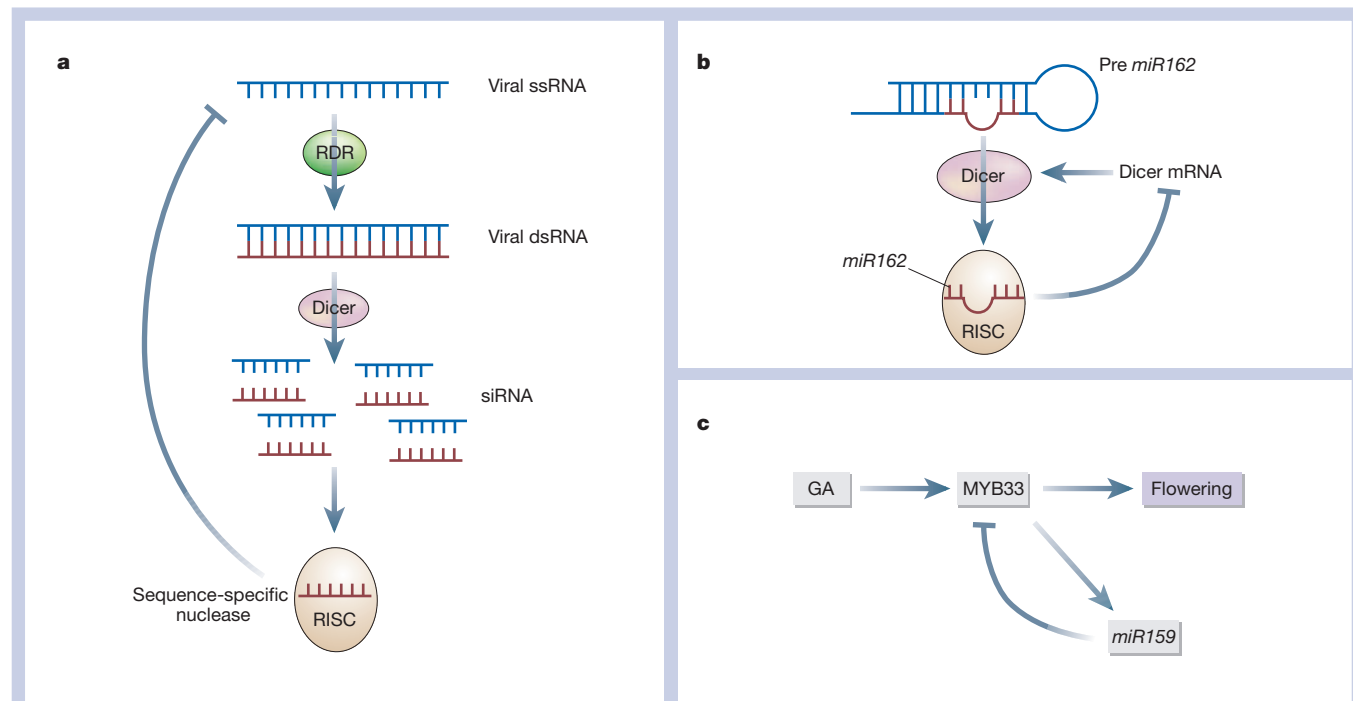


Figure 5 Feedback mechanisms in RNA silencing. **a**, The sequence of events when siRNA production involves an RDR using a ssRNA template. The siRNA is incorporated into RISC and negatively regulates its own production by targeting RISC at the ssRNA. **b**, The feedback inhibition of *miR162* on its target mRNA encoding DCL1 Dicer (ref. 38). DCL1 mediates the production of *miR162* from the pre-*miR162* precursor RNA. The *miR162* then targets the *DCL1* mRNA, and negatively regulates DCL1 synthesis. So a high level of *miR162* leads to a decrease in the rate of DCL1 production, whereas a low level of *miR162* has the opposite effect. **c**, A negative-feedback regulatory mechanism proposed for *miR159*. The target RNA of *miR159* encodes a MYB33 transcription factor. Both MYB33 mRNA and *miR159* are positively regulated by the plant hormone GA⁸⁵. In a GA-responding *A. thaliana* there is an increase in the level of the MYB33 mRNA which is associated with an increase in the level of *miR159*. The high levels of *miR159* then suppress the GA-stimulated increase in MYB33. Several rounds of the priming process would amplify the silencing effect of the siRNA.

(<http://cgrb.orst.edu/smallRNA/>) lists more than 2,000 of them but there are likely to be many more, and their extended analysis will probably be a rich source of information about genetic elements that are targeted by RNA silencing.

The RNA targeting mechanisms of RNA silencing have not been extensively investigated in plants: one particularly neglected area is the possibility of translational suppression, as is mediated by animal miRNAs⁵. One plant miRNA (*miR172*) is known to suppress translation of the floral regulator *AP2* mRNA^{35,86}, but for most other miRNAs and siRNAs the possibility of translational effects has not been investigated. Transgene silencing⁸⁷ suggests that translational effects may be more general, and perhaps protein rather than RNA profiling should be used to identify the targets of RNA silencing mechanisms.

Another outstanding question concerns silencing RNAs that move between cells. Current indications are that this signalling process affects virus movement⁵⁵ (E. Bayne, F. Schwach and D.B., unpublished work), but possible roles in plant growth and development have yet to be explored. *miR165* and *miR166* regulate spatial information in plant development^{34,67} and conceivably the movement of silencing-related RNAs is involved (Fig. 2).

Although RNA silencing is a relatively recent topic of research, the mechanisms are now becoming clear and there is some indication of its biological roles. The discovery of RNA silencing has completely changed our view of RNA as a regulatory molecule in eukaryotic cells and it is likely that this view will continue to evolve as further discoveries emerge about the diversity of silencing mechanisms. □

doi:10.1038/nature02874

- Wingard, S. A. Hosts and symptoms of ring spot, a virus disease of plants. *J. Agric. Res.* **37**, 127–153 (1928).
- Ratcliff, F., Harrison, B. D. & Baulcombe, D. C. A similarity between viral defense and gene silencing in plants. *Science* **276**, 1558–1560 (1997).
- Covey, S. N., Al-Kaff, N. S., Langara, A. & Turner, D. S. Plants combat infection by gene silencing. *Nature* **385**, 781–782 (1997).
- Hamilton, A. J. & Baulcombe, D. C. A species of small antisense RNA in post-transcriptional gene silencing in plants. *Science* **286**, 950–952 (1999).
- Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
- Wassenegger, M., Heimes, S., Riedel, L. & Sanger, H. L. RNA-directed *de novo* methylation of genomic sequences in plants. *Cell* **76**, 567–576 (1994).
- Mette, M. F., Aufsatz, W., van der Winden, J., Matzke, M. A. & Matzke, A. J. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J.* **19**, 5194–5201 (2000).
- Jones, L., Ratcliff, F. & Baulcombe, D. C. RNA-directed transcriptional gene silencing in plants can be inherited independently of the RNA trigger and requires Met1 for maintenance. *Curr. Biol.* **11**, 747–757 (2001).
- Zilberman, D., Cao, X. & Jacobsen, S. E. *ARGONAUTE4* control of locus specific siRNA accumulation and DNA and histone methylation. *Science* **299**, 716–719 (2003).
- Volpe, T. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833–1837 (2002).
- Lingel, A., Simon, B., Izaurralde, E. & Sattler, M. Structure and nucleic-acid binding of the *Drosophila* Argonaute2 PAZ domain. *Nature* **426**, 465–469 (2003).
- Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R. & Hannon, G. J. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**, 1146–1150 (2001).
- Liu, J. *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* published online 29 July 2004 (doi:10.1126/science.1102513).
- Vaucheret, H., Vazquez, F., Crete, P. & Bartel, D. P. The action of *ARGONAUTE1* in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. *Genes Dev.* **18**, 1187–1197 (2004).
- Verdel, A. *et al.* RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672–676 (2004).
- Taverna, S. D., Coyne, R. S. & Allis, D. C. Methylation of histone H3 at lysine 9 targets programmed DNA elimination in *Tetrahymena*. *Cell* **110**, 701–711 (2002).
- Mochizuki, K., Fine, N. A., Fujisawa, T. & Gorovsky, M. A. Analysis of a *piwi*-related gene implicates small RNAs in genome rearrangement in *Tetrahymena*. *Cell* **110**, 689–699 (2002).
- Hunter, C., Sun, H. & Poethig, R. S. The *Arabidopsis* heterochronic gene *ZIPPY* is an ARGONAUTE family member. *Curr. Biol.* **13**, 1734–1739 (2003).
- Fagard, M., Boutet, S., Morel, J.-B., Bellini, C. & Vaucheret, H. AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. *Proc. Natl Acad. Sci. USA* **97**, 11650–11654 (2000).
- Boutet, S. *et al.* *Arabidopsis* *HEN1*: A genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance. *Curr. Biol.* **13**, 843–848 (2003).
- Schauer, S. E., Jacobsen, S. E., Meinke, D. W. & Ray, A. *DICER-LIKE1*: blind men and elephants in *Arabidopsis* development. *Trends Plant Sci.* **7**, 487–491 (2002).
- Papp, I. *et al.* Evidence for nuclear processing of plant microRNA and short interfering RNA precursors. *Plant Physiol.* **132**, 1382–1390 (2003).
- Finnegan, E. J., Margis, R. & Waterhouse, P. M. Posttranscriptional gene silencing is not compromised in the *Arabidopsis* *CARPEL FACTORY* (*DICER-LIKE1*) mutant, a homolog of *Dicer-1* from *Drosophila*. *Curr. Biol.* **13**, 236–240 (2003).

- Xie, Z. *et al.* Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2**, E104 (2004).
- Hamilton, A. J., Voinnet, O., Chappell, L. & Baulcombe, D. C. Two classes of short interfering RNA in RNA silencing. *EMBO J.* **21**, 4671–4679 (2002).
- Tang, G., Reinhart, B. J., Bartel, D. P. & Zamore, P. D. A biochemical framework for RNA silencing in plants. *Genes Dev.* **17**, 49–63 (2002).
- Han, M.-H., Goud, S., Song, L. & Fedoroff, N. The *Arabidopsis* double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proc. Natl Acad. Sci. USA* **101**, 1093–1098 (2004).
- Vazquez, F., Gascoli, V., Crete, P. & Vaucheret, H. The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. *Curr. Biol.* **14**, 346–351 (2004).
- Rhoades, M. W. *et al.* Prediction of plant microRNA targets. *Cell* **110**, 513–520 (2002).
- Jones-Rhoades, M. W. & Bartel, D. P. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol. Cell* **14**, 787–799 (2004).
- Llave, C., Kasschau, K. D., Rector, M. A. & Carrington, J. C. Endogenous and silencing-associated small RNAs in plants. *Plant Cell* **14**, 1605–1619 (2002).
- Sunkar, R. & Zhu, J.-K. Novel and stress-regulated miRNAs and other small RNAs from *Arabidopsis*. *Plant Cell* (in the press).
- Palatnik, J. F. *et al.* Control of leaf morphogenesis by microRNAs. *Nature* **425**, 257–263 (2003).
- Emery, J., Floyd, S. K., Alvarez, J., Baum, S. F. & Bowman, J. L. Radial patterning of *Arabidopsis* shoots by Class III HD-ZIP and KANADI genes. *Curr. Biol.* **13**, 1768–1774 (2003).
- Chen, X. M. A microRNA as a translational repressor of *APETALA2* in *Arabidopsis* flower development. *Science* **303**, 2022–2025 (2004).
- Mallory, A. C., Dugas, D. V., Bartel, D. P. & Bartel, B. MicroRNA regulation of NAC-domain targets is required for proper formation and separation of adjacent embryonic, vegetative, and floral organs. *Curr. Biol.* **14**, 1035–1046 (2004).
- Llave, C., Xie, Z., Kasschau, K. D. & Carrington, J. C. Cleavage of *Scarecrow*-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* **297**, 2053–2056 (2002).
- Xie, Z., Kasschau, K. D. & Carrington, J. C. Negative feedback regulation of *Dicer-Like1* in *Arabidopsis* by microRNA-guided mRNA degradation. *Curr. Biol.* **13**, 784–789 (2003).
- Park, W., Li, J., Song, R., Messing, J. & Chen, X. CARPEL FACTORY: a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr. Biol.* **12**, 1484–1495 (2002).
- Floyd, S. K. & Bowman, J. L. Gene regulation: ancient microRNA target sequences in plants. *Nature* **428**, 485–486 (2004).
- Sijen, T. *et al.* On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* **107**, 465–476 (2001).
- Smardon, A. *et al.* EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in *C. elegans*. *Curr. Biol.* **10**, 169–178 (2000).
- Cogoni, C. & Macino, G. Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. *Nature* **399**, 166–169 (1999).
- Dalmay, T., Hamilton, A. J., Rudd, S., Angell, S. & Baulcombe, D. C. An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell* **101**, 543–553 (2000).
- Mourrain, P. *et al.* *Arabidopsis* *SGS2* and *SGS3* genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell* **101**, 533–542 (2000).
- Iyer, L. M., Koonin, E. V. & Aravind, L. Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. *BMC Struct. Biol.* **3**, 1–23 (2003).
- Dalmay, T. D., Hornefeld, R., Braunstein, T. H. & Baulcombe, D. C. *SDE3* encodes an RNA helicase required for post-transcriptional gene silencing in *Arabidopsis*. *EMBO J.* **20**, 2069–2078 (2001).
- Xie, Z., Fan, B., Chen, C. H. & Chen, Z. An important role of an inducible RNA-dependent RNA polymerase in plant antiviral defense. *Proc. Natl Acad. Sci. USA* **98**, 6516–6521 (2001).
- Makeyev, E. V. & Bamford, D. H. Cellular RNA-dependent RNA polymerase involved in posttranscriptional gene silencing has two distinct activity modes. *Mol. Cell* **10**, 1417–1427 (2002).
- Schiebel, W., Haas, B., Marinkovic, S., Klanner, A. & Sanger, H. L. RNA-directed RNA polymerase from tomato leaves. II. Catalytic *in vitro* properties. *J. Biol. Chem.* **268**, 11858–11867 (1993).
- Vaistij, F. E., Jones, L. & Baulcombe, D. C. Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative RNA-dependent RNA polymerase. *Plant Cell* **14**, 857–867 (2002).
- Voinnet, O., Vain, P., Angell, S. & Baulcombe, D. C. Systemic spread of sequence-specific transgene RNA degradation is initiated by localized introduction of ectopic promoterless DNA. *Cell* **95**, 177–187 (1998).
- Yamada, K. *et al.* Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science* **302**, 842–846 (2003).
- Martiniussen, R. Maintenance of heterochromatin by RNA interference of tandem repeats. *Nature Genet.* **35**, 1–2 (2003).
- Voinnet, O., Lederer, C. & Baulcombe, D. C. A viral movement protein prevents spread of the gene silencing signal in *Nicotiana benthamiana*. *Cell* **103**, 157–167 (2000).
- Palauqui, J.-C., Elmayan, T., Pollien, J.-M. & Vaucheret, H. Systemic acquired silencing: transgene-specific post-transcriptional silencing is transmitted by grafting from silenced stocks to non-silenced scions. *EMBO J.* **16**, 4738–4745 (1997).
- Voinnet, O. & Baulcombe, D. C. Systemic signalling in gene silencing. *Nature* **389**, 553 (1997).
- Timmons, L. & Fire, A. Specific interference by ingested dsRNA. *Nature* **395**, 854 (1998).
- Feinberg, E. H. & Hunter, C. P. Transport of dsRNA into cells by the transmembrane protein SID-1. *Science* **301**, 1545–1547 (2003).
- Winston, W. M., Molodowitch, C. & Hunter, C. P. Systemic RNAi in *C. elegans* requires the putative transmembrane protein SID-1. *Science* **295**, 2456–2459 (2002).
- Haywood, V., Kragler, F. & Lucas, W. J. Plasmodesmata: pathways for protein and ribonucleoprotein signaling. *Plant Cell* **14**, S303–S325 (2002).

62. Himber, C., Dunoyer, P., Moissiard, G., Ritzenthaler, C. & Voinnet, O. Transitivity-dependent and -independent cell-to-cell movement of RNA silencing. *EMBO J.* **22**, 4523–4533 (2003).
63. Lakatos, L., Szitty, G., Silhavy, D. & Burgyan, J. Molecular mechanism of RNA silencing suppression mediated by p19 protein of tombusviruses. *EMBO J.* **23**, 876–884 (2004).
64. Mallory, A. C. *et al.* HC-Pro suppression of transgene silencing eliminates the small RNAs but not transgene methylation or the mobile signal. *Plant Cell* **13**, 571–583 (2001).
65. Ding, B., Kwon, M.-O., Hammond, R. & Owens, R. Cell-to-cell movement of potato spindle tuber viroid. *Plant J.* **12**, 931–936 (1997).
66. Kim, M., Canio, W., Kessler, S. & Sinha, N. Developmental changes due to long-distance movement of a homeobox fusion transcript in tomato. *Science* **293**, 287–289 (2001).
67. Juarez, M. T., Kul, J. S., Thomas, J., Heller, B. A. & Timmermans, M. C. microRNA-mediated repression of *rolled leaf1* specifies maize leaf polarity. *Nature* **428**, 84–88 (2004).
68. Kidner, C. A. & Martienssen, R. Spatially restricted microRNA directs leaf polarity through *ARGONAUTE1*. *Nature* **428**, 81–84 (2004).
69. Yoo, B.-C. *et al.* A systemic small RNA signaling system in plants. *Plant Cell*, 1979–2000 (2004).
70. Moissiard, G. & Voinnet, O. Viral suppression of RNA silencing in plants. *Mol. Plant Pathol.* **5**, 71–82 (2004).
71. Szitty, G., Molnar, A., Silhavy, D., Hornyik, C. & Burgyan, J. Short defective interfering RNAs of tombusviruses are not targeted but trigger post-transcriptional gene silencing against their helper virus. *Plant Cell* **14**, 359–372 (2002).
72. Wang, M.-B. *et al.* On the role of RNA silencing in the pathogenicity and evolution of viroids and viral satellites. *Proc. Natl Acad. Sci. USA* **101**, 3275–3280 (2004).
73. Chapman, E. J., Prokhnovsky, A. I., Gopinath, K., Dolja, V. & Carrington, J. C. Viral RNA silencing suppressors inhibit the microRNA pathway at an intermediate step. *Genes Dev.* **18**, 1179–1186 (2004).
74. Dunoyer, P., Lecellier, C. H., Parizotto, E. A., Himber, C. & Voinnet, O. Probing the microRNA and small interfering RNA pathways with virus-encoded suppressors of RNA silencing. *Plant Cell* **16**, 1235–1250 (2004).
75. Mallory, A. C., Reinhart, B. J., Bartel, D., Vance, V. B. & Bowman, L. H. A viral suppressor of RNA silencing differentially regulates the accumulation of short interfering RNAs and microRNAs in tobacco. *Proc. Natl Acad. Sci. USA* **99**, 15228–15233 (2002).
76. Vargason, J. M., Szitty, G., Burgyan, J. & Tanaka Hall, T. M. Size selective recognition of siRNA by an RNA silencing suppressor. *Cell* **115**, 799–811 (2003).
77. Ye, K., Malinin, L. & Patel, D. J. Recognition of small interfering RNA by a viral suppressor of RNA silencing. *Nature* **426**, 874–878 (2003).
78. Li, W.-X. *et al.* Interferon antagonist proteins of influenza and vaccinia virus are suppressors of RNA silencing. *Proc. Natl Acad. Sci. USA* **101**, 1350–1355 (2004).
79. Schwarz, D. S. *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199–208 (2003).
80. Li, H., Li, W. X. & Ding, S. W. Induction and suppression of RNA silencing by an animal virus. *Science* **296**, 1319–1321 (2002).
81. Pfeffer, S. *et al.* Identification of virus-encoded microRNAs. *Science* **304**, 734–736 (2004).
82. Lichner, Z., Silhavy, D. & Burgyan, J. Double-stranded RNA-binding proteins could suppress RNA interference-mediated antiviral defences. *J. Gen. Virol.* **84**, 975–980 (2003).
83. Kasschau, K. D. *et al.* P1/HC-Pro, a viral suppressor of RNA silencing, interferes with *Arabidopsis* development and miRNA function. *Dev. Cell* **4**, 205–217 (2003).
84. Pruss, G. J. *et al.* The potyviral suppressor of RNA silencing confers enhanced resistance to multiple pathogens. *Virology* **320**, 107–120 (2004).
85. Achard, P., Herr, A. J., Baulcombe, D. & Harberd, N. P. Modulation of photoperiodic control of floral transition by a hormonally regulated microRNA. *Genes Dev.* (in the press).
86. Aukerman, M. J. & Sakai, H. Regulation of flowering time and floral organ identity by a microRNA and its *APETALA2-like* target genes. *Plant Cell* **15**, 2730–2741 (2003).
87. VanHoudt, H., Ingelbrecht, L., VanMontagu, M. & Depicker, A. Post-transcriptional silencing of a neomycin phosphotransferase II transgene correlates with the accumulation of unproductive RNAs and with increased cytosine methylation of 3' flanking regions. *Plant J.* **12**, 379–392 (1997).

Acknowledgements I thank the Gatsby Charitable Foundation for supporting work in my laboratory.

Competing interests statement The author declares that he has no competing financial interests.

The role of RNA interference in heterochromatic silencing

Zachary Lippman & Rob Martienssen

Watson School of Biological Sciences and Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA
(e-mail: martienss@cshl.edu)

Soon after its discovery 75 years ago, heterochromatin, a dense chromosomal material, was found to silence genes. But its importance in regulating gene expression was controversial. Long thought to be inert, heterochromatin is now known to give rise to small RNAs, which, by means of RNA interference, direct the modification of proteins and DNA in heterochromatic repeats and transposable elements. Heterochromatin has thus emerged as a key factor in epigenetic regulation of gene expression, chromosome behaviour and evolution.

The botanist Emil Heitz first defined heterochromatin as nuclear material that remains condensed throughout the cell cycle, unlike the rest of the chromosome, which unravels between cell divisions (Fig. 1)¹. Biology textbooks now portray heterochromatin as a 'junkyard' of silent noncoding DNA and defective transposons. This picture has emerged because heterochromatin is composed of DNA sequences with little or no coding potential, repeated thousands of times, and silenced by the covalent modification of the DNA itself and of the histones around which the DNA is wound².

But heterochromatin has a controversial past because of its ability to influence the regulation of nearby genes. In *Drosophila*, juxtaposition of eye-colour genes with heterochromatin results in eyes that are mottled red and white (Fig. 1). This observation led Muller to coin the term 'position effect variegation' (PEV)^{2,3}. Indeed, Goldschmidt working on PEV in *Drosophila*, and McClintock working with 'controlling elements' or transposable elements in maize (Fig. 1), sought to elevate the status of heterochromatic gene silencing to that of a regulatory mechanism underlying development — an idea that received little support at the time⁴. Recently, however, as the genetic basis for PEV, transgene silencing, viral resistance and transposon regulation has emerged, along with the sequence of heterochromatic regions in plant and animal genomes, this idea has gained credence. One such regulatory mechanism relies on post-transcriptional regulation mediated by RNA, or RNA interference (RNAi). Unexpectedly, RNAi has been found to have a central role in heterochromatic gene silencing, despite the classical view that 'silent' heterochromatin is not transcribed into RNA.

Here, we review how heterochromatic silencing depends on the processing of repeat RNA transcripts into short interfering RNAs (siRNAs), which then direct chromatin modification. This mechanism explains how different repeats found in various eukaryotic genomes can be similarly incorporated into heterochromatin. Although many questions remain, these studies have resurrected the controversial suggestion that heterochromatic silencing is important in evolution and development.

Repeated sequences and their modification

Most heterochromatin is found near centromeres and telomeres, and consists of tandem (satellite) repeats, which are sometimes interrupted by transposable elements (Box 1). Pericentromeric repeats range from a few kilobases (kb) in

fission yeast to 100–400-base pair (bp) short repeats that are arranged in Megabase-pair (Mb) arrays in mammals, plants and *Drosophila* (Box 1). Heterochromatic repeats share little similarity between species, but in all cases, heterochromatin is silenced by conserved epigenetic modifications of histones and DNA. This epigenetic silencing, as well as the higher-order packaging of repeats into heterochromatin, is believed to prevent illegitimate recombination, which can lead to chromosomal rearrangements. Furthermore, chromosomes are protected from active transposons, which can cause mutations when they are integrated into genes.

Methylation, acetylation, phosphorylation and ubiquitination of the core histones H2A, H2B, H3 and H4, and histone variants such as H2A.Z and H3.3, are implicated in gene regulation. These modifications are collectively referred to as the histone code⁵. Many of these modifications are specific for heterochromatin or euchromatin, such as methylation of histone H3 lysine 9 and lysine 4, respectively. In *Drosophila*, the suppressor of PEV *Su(var)3-9* (*Suppressor of variegation 3-9*) encodes a SET (*Su(var)3-9, Enhancer-of-zeste, Trithorax*)-domain protein that is conserved in plants, animals and yeast. This protein is responsible for histone H3 lysine 9 methylation (H3mK9)⁶. Methylated lysine residues on histone H3 are recognized by chromo-domain proteins such as the highly conserved heterochromatin protein 1 (HP1)⁶, which is also a suppressor of PEV. The crystal structures of SET domains and chromo-domains indicate specific residues that determine which methylated lysines are recognized by each protein⁷.

The role of DNA methylation in heterochromatic gene silencing was recognized before that of histone modification⁸ (Box 2), even though it is less well conserved. DNA methylation is absent, or nearly absent in yeast, flies and nematodes, but a link between DNA methylation and histone methylation is well established in fungi apart from yeast, and in animals and plants⁸. Therefore, it is likely that DNA and histone modifications have a common role in gene silencing: they may even have a common origin.

In *Neurospora crassa*, DNA methylation depends on the *Su(var)3-9* lysine 9 methyltransferase *Dim-5* (ref. 9), whereas in *Arabidopsis*, plant-specific CNG and CNN methylation (where N is A, C, G or T) depends on the *Su(var)3-9* homologue KRYPTONITE (*KYP*; also known as *SUVH4*), and on the CNG and CNN DNA methyltransferase *CHROMOMETHYLASE3* (*CMT3*)^{10–13}. In *Arabidopsis*, methylation at histone H3 lysine 9 is lost from many transposons in the CG DNA methyltransferase mutant *met1* (ref. 14), but it is

retained by others¹⁵, and by centromeric satellite repeats¹⁶. In mammals, the DNA-methyltransferases Dnmt1 and Dnmt3b interact with the Suv39h (Suppressor of variegation 3-9 homologue) H3K9 methyltransferase and HP1, respectively, providing biochemical support for a functional link between DNA and histone methylation^{17,18}. Loss of H3mK9 in *suvh39*^{-/-} knockout embryonic stem cells reduces DNA methylation at major centromeric satellites, but has little impact elsewhere¹⁹. Although *dnmt1*^{-/-} or *dnmt3*^{-/-} cells have normal amounts of heterochromatic H3mK9 (ref. 19), H3mK9 is lost when cells are treated with DNA methylation inhibitors²⁰. So, histone H3 lysine 9 methylation and DNA methylation are interdependent, perhaps as a result of interactions between their respective methyltransferases. However, the inter-relationship between histone and DNA modification, although evolutionarily conserved, must depend, at least in part, on sequence specificity because not all DNA sequences are targeted equally for these modifications.

In *Arabidopsis*, the chromatin remodelling ATPase DDM1 (decrease in DNA methylation)²¹ is required for both DNA methylation and H3mK9 (refs 16, 22): *ddm1* mutants have been isolated along with *met1* mutants in screens for loss of centromeric-repeat methylation^{23,24}. Nearly all the targets of *DDM1* are transposons and tandem repeats, which are the focus of both DNA and histone methylation in plant genomes^{22,25,26}. Both CG and CNG DNA methyltransferases contribute to transposon silencing in *Arabidopsis*²⁷, as does histone deacetylation^{15,28}. H3mK9 seems to have a relatively minor role in transposon silencing, although double mutants between *kyp* and other histone methyltransferases have not yet been examined. As almost all DNA methylation and H3mK9 is

confined to transposons and repeats, these elements must somehow be distinguished from genes. RNAi is one mechanism by which sequence-specific targeting might be achieved.

RNAi-mediated heterochromatic modifications in yeast

Heterochromatic silencing in the fission yeast *Schizosaccharomyces pombe* occurs at the centromeres, telomeres and the mating-type loci. Pericentromeric heterochromatin is composed of complex repeats, known as *dg* and *dh*. A single copy of each repeat is also found at the mating-type locus (Box 1). Silencing depends on the histone deacetylases Clr3 and Clr6, the histone H3K9 methyltransferase Clr4, and on the HP1 homologue Swi6 (ref. 29).

Key components of the RNAi pathway include a small RNA-binding protein called Argonaute (Ago), a dsRNA ribonuclease termed Dicer (DCR), and an RNA-dependent RNA polymerase (RdRP; also known as RDR) (see reviews in this issue by Meister and Tuschl, page 343, and Baulcombe, page 356). In *S. pombe*, unlike higher eukaryotes, each of these genes is unique. *ago*⁻, *dcr*⁻ and *rdp*⁻ mutants are viable but defective in silencing of reporter genes integrated into centromeric (but not mating-type) repeats³⁰. Transcripts from both strands of the centromeric repeats are the targets of RNAi³⁰, and siRNAs corresponding to these repeats can be detected³¹. In all three RNAi mutants, repeat-associated H3mK9 is reduced, and both H3mK9 and Swi6 are lost from centromeric reporter genes, demonstrating that RNAi is responsible for the heterochromatic modifications³⁰. H3mK9 recruits Swi6 by means of its highly conserved chromodomain⁷, which in turn recruits cohesin — a highly conserved protein required for sister-chromatid cohesion and chromosome segregation during mitosis³². Loss of cohesin results in lagging chromosomes at anaphase in each of the RNAi mutants, as it does in *clr4*⁻ and *swi6*⁻ mutants^{33,34}.

Biochemical purification of chromodomain complexes in fission yeast has yielded the RITS (RNAi-induced transcriptional gene silencing) complex, which includes siRNA from centromeric repeats as well as Ago, a chromodomain protein Chp1, and a novel protein Tas3 (ref. 35) (Fig. 2a). Chp1 and Tas3 are bound to the centromeres in an RNAi-dependent manner, and in this respect they resemble RdRP³⁰, which may be recruited to the centromere by means of siRNAs and nascent repeat transcripts³⁶. It is possible that siRNA also recruits the RITS complex by means of Ago³⁵, whose PAZ domain (named after piwi-argonaute-zwille) binds siRNA (see reviews in this issue by Meister and Tuschl, page 343, and Ambros, page 350). The RITS complex may also be recruited to the centromere by means of the Chp1 chromodomain, which binds H3mK9 (refs 35, 37). This would explain the reliance of RNAi on the H3K9 methyltransferase Clr4 (ref. 38). In either case, the mechanism by which H3mK9 arrives at centromeric repeats in the first place is not clear (Fig. 2a).

Mating-type silencing is unaffected in *ago*⁻, *dcr*⁻ and *rdp*⁻ mutants³⁰, as well as in *chp1*⁻ and *tas3*⁻ mutants³⁵, indicating that some other mechanism must be responsible for the maintenance of mating-type silencing in the absence of RNAi. This mechanism involves histone deacetylation because silencing is lost in histone deacetylase mutants, or when histone deacetylase is inhibited by trichostatin A (TSA)²⁹. When TSA inhibition is removed, mating-type silencing only returns slowly in RNAi mutants³⁹. Unstable transgene silencing conferred by ectopic mating-type repeats also depends on RNAi^{33,39}. These observations suggest a role for RNAi in establishment rather than in maintenance of silencing³⁹. However, it has recently been reported that maintenance of mating-type silencing also requires RNAi, but only in the absence of either one of two transcription factors (presumably repressors) in the ATF/CREB (activating transcription factor/cAMP response-element binding protein) family⁴⁰. So RNAi has a role in maintenance of mating-type silencing, which is similar to its role at the centromere³⁰, but this role is obscured by ATF.

The recent excitement surrounding the role of RNAi and microRNAs (miRNAs) in development (see review in this issue by Ambros, page 350) has overshadowed the fact that the first biological

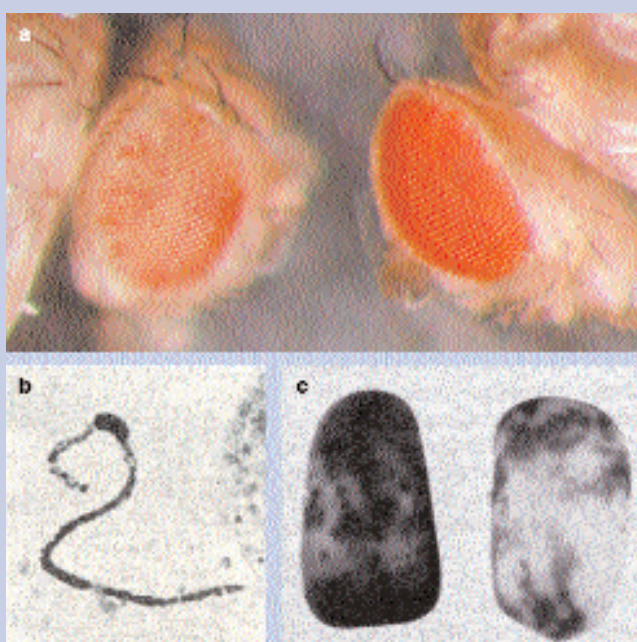
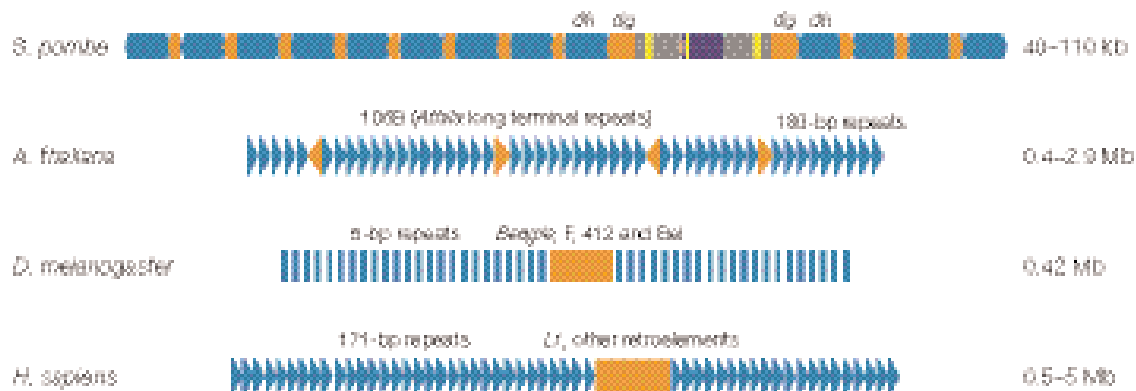


Figure 1 Heterochromatin, transposable elements and PEV. **a**, PEV in the *Drosophila* eye results in red-white mottling (left). It is caused by the juxtaposition of heterochromatin with the *white* locus following inversion (*mottled-w4*). PEV is suppressed by the RNA-helicase mutant *spindle-E/homeless* (right) and other mutants in RNAi (reproduced with permission from ref. 62). **b**, Heterochromatic knobs on maize chromosome 7 stained at pachytene. Variability in the size and location of knobs was used by B. McClintock to trace the origin of maize landraces in central America (reproduced with permission from ref. 83). **c**, Transposon-mediated variegation in maize. Defective derivatives of the CACTA transposable element, *dSpm* (or inhibitor, *I*), can epigenetically control the anthocyanin gene *A1* when they are integrated into the *A1* promoter. In the absence of the autonomous controlling element *Suppressor-mutator* (*Spm*), or *Enhancer* (*En*), gene expression is lost during development (right) leading to variegation (reproduced with permission from Cold Spring Harbor Archives).

Box 1

Heterochromatic repeats



Heterochromatic repeats are typically long arrays of short DNA sequences arranged in tandem head-to-tail orientation, and are concentrated in centromeric regions. Although unrelated in sequence, these arrays are found in plants, animals and fungi and are often interrupted by transposable elements. Schematic representations of pericentromeric heterochromatin in fission yeast²⁹, *Drosophila*⁸⁸, *Arabidopsis*⁸⁹ and humans⁹⁰ are shown. Pericentromeric heterochromatin is composed of tandem satellite repeats (blue) interrupted by retrotransposable elements (orange). Names for each repeat are given. The size of satellite repeat arrays is indicated on the right. Fission yeast repeats are relatively long (a few kb), whereas *Drosophila* repeats are very short (a few bp) and are interrupted by specific classes of retrotransposable element, such as the *Ty3/gypsy*-class elements *HMS Beagle* and *412*. Plants and mammals have short

repeats (170–300 bp) and are also interrupted by retrotransposons.

In plants, pericentromeric regions contain many *gypsy*-class retrotransposons, but few other types⁸⁹. Knobs are also interrupted by retrotransposons⁹¹. Knob-like structures in maize vary widely between natural ecotypes⁸³, as they do in *Arabidopsis*⁸⁴. In ecotype Columbia, the knob on the short arm of chromosome 4 (*hk4S*) has become a model for functional studies of heterochromatin because its complete sequence is known²⁶. A 1.95-kb tandem repeat represented 22.5 times is surrounded by more than 30 DNA transposons and 40 retrotransposons with few interspersed genes. The interspersed gene islands with retrotransposon clusters throughout the maize genome resembles the *Arabidopsis* knob²⁵, and a similar organization has recently been reported for a centromere in rice⁹².

function attributed to RNAi was defence against transposons and viruses (see review in this issue by Baulcombe, page 356). Many of the genes required for RNAi in *Caenorhabditis elegans*, *Drosophila* and *Chlamydomonas* are also required for transposon suppression^{41–43}. Moreover, siRNAs cloned from *Drosophila* and trypanosomes match several classes of repetitive sequences^{44,45}. The ability of *Drosophila* heterochromatic tandem repeats, known as *Suppressor of Stellate* (*Su(Ste)*), to silence the homologous *Stellate* genes depends on the RNA helicase *spindle-E* (*spn-E*)⁴³. Mutations in this gene lead to the activation of transposons and other genomic repeats. In a dramatic example from *Tetrahymena*, macronuclear-DNA-elimination of repeats is RNAi-dependent and requires H3mK9, the AGO homologue *TW11* (related to *piwi*) and a chromodomain protein Pdd1p (programmed DNA degradation) (refs 46, 47).

Heterochromatic silencing in plants

Heterochromatic silencing in plants also involves RNAi, but DNA methylation is an additional factor (Box 2). Most (90–95%) endogenous siRNAs in *Arabidopsis* correspond to transposons and repeats whose histones and DNA are heavily methylated^{16,22,26,48}. Readthrough transcription of inverted repeats, transcription after the insertion of repeats into another transposon, or other mechanisms could account for the occurrence of transposon dsRNA⁴⁹. Moreover, bidirectional transcription has been detected in *met1* and *ddm1* mutants²⁶. As long as they are transcribed from one strand, tandem repeats can theoretically maintain a population of siRNAs through multiple rounds of replication mediated by RdRP³⁶.

Despite the prevalence of siRNAs in *Arabidopsis*, only a handful of *Arabidopsis* transposons are activated in RNAi mutants^{15,48,50}. AGO and other RNAi components are encoded by partially redundant multigene families (see review in this issue by Baulcombe, page 356),

and at least some double *ago* mutants are lethal, complicating this analysis⁵¹. However RNAi-mediated silencing cannot be the only mechanism by which transposons are silenced, because siRNAs for many transposons are lost in *dicer-like3* (*dcl3*) and *rna-dependent rna polymerase2* (*rdp2*) mutants, but silencing is hardly affected⁴⁸.

Once transposon silencing has been established by RNAi, silencing may be maintained by other means. In *met1* and *ddm1* mutants, which lose DNA methylation, transposons are activated, and most remain active (preset) in backcrosses^{15,26}. This means that MET1 and DDM1 can maintain transposon silencing, but cannot re-establish it once it is lost. Transposon silencing is also lost in the histone deacetylase mutant *hda6/sil1* (where *sil1* is an allele of *hda6*) but, in this case, silencing is re-established in backcrosses to wild-type plants. Most transposon siRNAs are lost from *met1* mutants, but are retained in *sil1* mutants. Furthermore, MET1 can re-silence some transposons in backcrosses, and the siRNA corresponding to these transposons are retained in *met1* mutants¹⁵. It is possible, therefore, that siRNA (*in cis*) may be required for silencing *de novo* by means of MET1 and SIL1. A silencing complex that includes SIL1, MET1, DDM1 and RNA has been proposed on the basis of parallel complexes found in mammalian cells (Fig. 2b)¹⁵.

Although most silencing targets are transposons, phenotypic change can result from the silencing of genes. For example, the imprinted gene *FWA* encodes a homeodomain protein that controls flowering, and it is normally expressed in the endosperm; however, *FWA* is silenced in vegetative tissues by methylation of tandem repeats in its promoter^{52,53}. *FWA* transgenes are silenced when they are integrated into wild-type plants by means of *Agrobacterium* transformation, but not when they are transformed into *dcl3*, *rdp2* or *ago4* mutants, in which case ectopic expression and late flowering results⁵⁴. Double mutants of the redundant *de novo* non-CG DNA

methyltransferases domains rearranged methylase 1 and domains rearranged methylase 2 (*drm1* and *drm2*) (Box 2) also fail to silence introduced transgenes: thus DRM1 and DRM2 are likely to establish DNA methylation in this case^{54–56}. RNAi mutants (and loss of non-CG methylation) have little impact on silencing of endogenous *FWA*⁵⁴, but CG methylation and silencing is lost in *met1* mutants, resulting in heritable late-flowering mutant epialleles⁵². This indicates a role for *MET1* (Dnmt1-like) in maintenance of silencing, and a role for RNAi, *DRM1* and *DRM2* in establishment of silencing (Fig. 2b)⁵⁴. However, the possibility that RNAi has a redundant role in silencing maintenance cannot be ruled out.

Interestingly the tandem repeats that make up the promoter of *FWA* are generated by the insertion of a short interspersed nucleotide element (SINE) related to *A. thaliana* SINE 2 (*AtSN2*) (ref. 26). These repeats accumulate siRNA and H3mK9, both of which are lost in *met1* mutants (ref. 26). So loss of transposon siRNA may account for the inheritance of late-flowering *FWA* epialleles from *met1* mutants, and for the failure to establish *FWA*-transgene silencing in *dcl3*, *rdr2* and *ago4* mutants⁵⁴. siRNAs corresponding to *AtSN1* are also lost in all four mutants^{15,48,50}, indicating that SINE elements may be targets of this silencing mechanism.

The loss of siRNA from *met1* and *ddm1* mutants and the epigenetic inheritance of transposon activity in plants may be similar to the loss of siRNA in *clr4*[–] mutants and the epigenetic inheritance of mating-type de-repression in *S. pombe*^{38,39}. In both cases, ‘maintenance’ chromatin-modification enzymes may also function in establishment by guiding interactions with siRNAs. Despite differences in DNA methylation, the dependence of heterochromatin silencing on RNAi is similar in *S. pombe* and *Arabidopsis*, including the strand-specific regulation of centromeric repeats (B. May *et al.*, unpublished data). Furthermore, fission yeast retrotransposon silencing is more dependent on histone deacetylation than on RNAi (K. R. Hansen *et al.*, unpublished data), resembling *Arabidopsis* transposon silencing in this respect (ref. 15). Therefore, both DNA methyltransferases and histone deacetylases on the one hand, and RNAi and histone methyltransferases on the other, participate in establishment and maintenance of heterochromatic silencing, although different targets may be involved in each mechanism^{15,28,30}.

RNAi-dependent silencing in plants

The earliest examples of sequence-specific RNA-mediated gene silencing came from plants (Box 3). Recently, models for RNA-dependent silencing have allowed the isolation of silencing mutants in *Arabidopsis*. Transcriptional gene silencing (TGS) and methylation of three different genes depends, at least in part, on the transcription of a homologous inverted repeat, which ultimately results in RNA-dependent DNA methylation or RdDM. Silencing of *PAI2* and *SUPERMAN* (*SUP*) is stabilized by inverted repeats of each entire gene, including the promoter^{10,55,57}. In contrast, silencing of a bacterial transgene can be achieved using inverted repeats of the nopaline synthase promoter (NOSpro) alone⁵⁸. In at least two of these three examples, transcription of the inverted repeat is required for silencing^{28,59}. Furthermore, for all these genes, histone and DNA methylation is disrupted in otherwise viable and fertile mutants. Silencing of *PAI2* and *SUP* requires *CMT3* and *KYP*^{10,11,13}, whereas silencing of NOSpro depends on *MET1* and on the histone deacetylase *HDA6/SIL1* (ref. 28).

Only one in more than forty silencing mutants isolated in the three screens for disrupted silencing of *PAI2*, *SUP* or *NOSpro*, is defective in RNAi: a single mutant allele of the Ago homologue *ago4* is defective in silencing *SUP*⁵⁰. It might, therefore, be concluded that RNAi is of little importance in RNA-dependent DNA methylation. However, at least one of the inverted-repeat transgenes (NOSpro) gives rise to siRNA, suggesting that the RNAi machinery is involved²⁸. Furthermore, inverted-repeat silencing systems in addition to *SUP* have defects in non-CG methylation in *ago4* mutants, resembling *drm1* and *drm2* double mutants⁵⁵, even when silencing of target loci is

Box 2

DNA methylation and histone modification

Transposons and repeats are heavily methylated and are associated with histone H3 lysine 9 methylation in mammals, plants and fungi (see text). Immunocytological staining has revealed that these modifications are focussed on heterochromatin where these sequences reside. Genes are almost free of these modifications in plants and filamentous fungi, but most exons are methylated in mammals^{8,25,93}. DNA methylation is less well conserved than histone methylation: the major cytosine 5-methyltransferase found in mammals (Dnmt1) is also found in plants (*MET1*), but is absent from *Drosophila*, *S. cerevisiae*, *S. pombe* and *C. elegans* in which DNA methylation is absent or nearly so. This suggests that it has been lost from many animal and fungal lineages²⁵. Dnmt1 prefers a symmetric CG context and favours hemimethylated DNA, thereby maintaining symmetric CG methylation patterns. The *de novo* methyltransferase Dnmt3 has a similar pattern of conservation, but can methylate a broader range of cytosines than Dnmt1. Plants have two redundant homologues of Dnmt3 (known as DRM1 and DRM2) but these proteins differ from animals in that they methylate cytosines in CNG and in asymmetric (CNN) contexts. Plants also are unique in having chromodomain-containing methyltransferases (known as CMT1, CMT2 and CMT3), at least one of which (CMT3) functions to maintain primarily CNG methylation²⁸.

unaffected⁶⁰. Moreover, the inverted repeats themselves are hypomethylated in *ago4* (ref. 60), which may contribute to production of functional protein in the case of *SUP*. Finally, a mutant allele of another Ago homologue, *ago1*, has defects in DNA methylation that are associated with post-transcriptional gene silencing (PTGS)⁶¹. Thus RNAi clearly plays a role in RdDM in plants, but this role can be obscured by other silencing mechanisms.

Transposons are regulated by the same mechanisms as those required for RdDM¹⁵. Most retrotransposons are silenced by *MET1* and *HDAC6/SIL1*; the retrotransposons *AtCOPIA44/TA3* and *AtCOPIA4* also require the H3K9 methyltransferase *KYP*, the chromomethylase *CMT3* and either *AGO4* or *AGO1* (refs 15, 50) (Fig. 2b). It is possible that *SUP*, *PAI2* and the transposable element *COPIA* on the one hand, and *NOSpro* and the transposable element *GYPSY* on the other, have distinct classes of *cis*-acting regulatory sequences. Alternatively, *SUP*, *PAI2* and *COPIA* may represent unstable (facultative) heterochromatin that requires H3mK9 for silencing. *NOSpro* and *GYPSY* may be stably silenced by further histone deacetylation (such as that of H2A, H2B and H4), as well as by DNA methylation. Further analysis of histone modifications will be required to identify which of these possibilities is the case.

RNAi and heterochromatin in animals

RNAi also has a role in heterochromatin formation in animals. In *Drosophila*, the Ago mutants *aubergine* and *piwi* disrupt TGS of transgene tandem repeats, which lose H3mK9 (ref. 62). In *spindle-E* mutants, H3mK9 is also lost, and two HP1 homologues are no longer associated with heterochromatic foci⁶². Like the histone H3K9 methyltransferase *Su(var)3-9*, all these mutants also suppress PEV⁶². These same genes regulate the post-transcriptional silencing of retrotransposons and tandem repeats, in addition to aspects of transcriptional silencing of transgenes⁶³. Also in *Drosophila*, mutants of *dcr-2*, but not *dcr-1*, affect the accumulation of siRNA derived from an inverted repeat transgene⁶⁴. *Drosophila dcr-2* mutants are similar to *dcl3* mutants in *Arabidopsis* in that both primarily affect siRNA, and are phenotypically wild type in all other respects. This suggests that in *Drosophila*, DCR-2 is only involved in the silencing of heterochromatin. *aubergine* and *piwi* mutants, on the other hand, resemble *Arabidopsis ago1* mutants in having defects

in transgene and heterochromatic silencing, and in having strong developmental phenotypes^{61,63}.

A role for RNAi in centromeric silencing in *C. elegans* is difficult to assess because canonical centromeres are lacking. However, RNAi regulates transposons in the germ line, although not in somatic cells in which Tc1 transposable elements are active⁴⁹. Silencing of extra-chromosomal transgene arrays also occurs in the germ line and is associated with H3mK9 (ref. 65). So, germ line heterochromatin formation may depend on RNAi. It would not be surprising if transposons were also associated with H3mK9. Germline-specific components of the RNAi machinery have been identified in *C. elegans*⁴⁹, which may be directly involved in targeting heterochromatin formation as occurs in *S. pombe*, *Arabidopsis* and *Drosophila*.

Linking RNAi with heterochromatic silencing in mammals has been more difficult, particularly because very few repeat-derived siRNAs have been cloned (see review in this issue by Ambros, page 350). Nonetheless, the localization of H3mK9 and HP1 to pericentromeric heterochromatin in mouse is abolished by RNase treatment⁶⁶. Moreover, mice with null mutations of the *Suv39h* histone methyltransferases show reduced viability and chromosomal instabilities¹⁹, reminiscent of *clr4*⁻ mutants in *S. pombe*, which are also defective in RNAi³⁸. *Dicer*-null mice arrest very early in development, and this has been attributed to a defect in stem-cell maintenance⁶⁷. Lethality could also result from the loss of centromeric silencing and chromosome segregation defects, which may be more problematic for multicellular organisms than for *S. pombe*. Centromeric satellite repeats are methylated in wild-type mouse cells, and some classes of repeats are expressed from both strands^{19,68}. The generation of conditional *Dicer* knockouts will be necessary to test for transcriptional activation of these repeats and for changes in methylation. Finally, two out of nine *Ago* homologues in mice have been knocked out, and both mutants show spermatogenesis defects and male sterility^{69,70}. One of these mutants is blocked completely at meiosis I (ref. 70), which could be a consequence of heterochromatin defects.

RNAi-dependent heterochromatin and chromosomal imprinting

At least in principle, RNAi-dependent histone and DNA modification can 'imprint' chromosomal sequences in such a way that these modifications are subsequently maintained throughout cell division — resembling heterochromatin in this respect¹. This might, in part, account for epigenetic mechanisms of gene silencing in plants and animals, such as imprinting (whereby gene expression depends on the parent of origin for the chromosome). Parent-of-origin-specific imprints imposed in the germ line are 'remembered' after fertilization, and are associated with changes in DNA and histone H3 methylation^{71,72}. In plants, both *FWA* and *MEDEA* are imprinted in the endosperm, and are associated with tandem repeats and DNA methylation^{53,55,73,74}. The silencing of *FWA* depends on RNAi and heterochromatic transposons^{26,54}, although a role for RNAi in the gametophyte has not yet been demonstrated. Transposons and repeats have been found at many imprinted loci in mammals and noncoding transcripts have been implicated. However, a role for RNAi itself is far from clear^{71,75,76}.

Paramutation in maize is an example of a chromosomal imprint where a silent state can be transferred from one allele to another in somatic cells⁷⁷. This imprint is stable and heritable, and depends on upstream tandem repeats that are differentially methylated in paramutagenic and paramutable alleles⁷⁷. It will be interesting to determine whether these repeats correspond to small RNAs.

RNA components direct several other forms of epigenetic silencing, and although a role for RNAi has not been demonstrated, aspects of this heterochromatic regulation share common features with that mediated by RNAi. X-chromosome inactivation in mammals is responsible for dosage compensation in XX females⁷⁸, and is mediated by a noncoding RNA, *Xist* (X-inactive specific transcript). Transcripts from both strands of this internally repetitive region are required for X-chromosome silencing, which is characterized by histone and DNA methylation⁷⁹. X-chromosome inactivation is imprinted in the

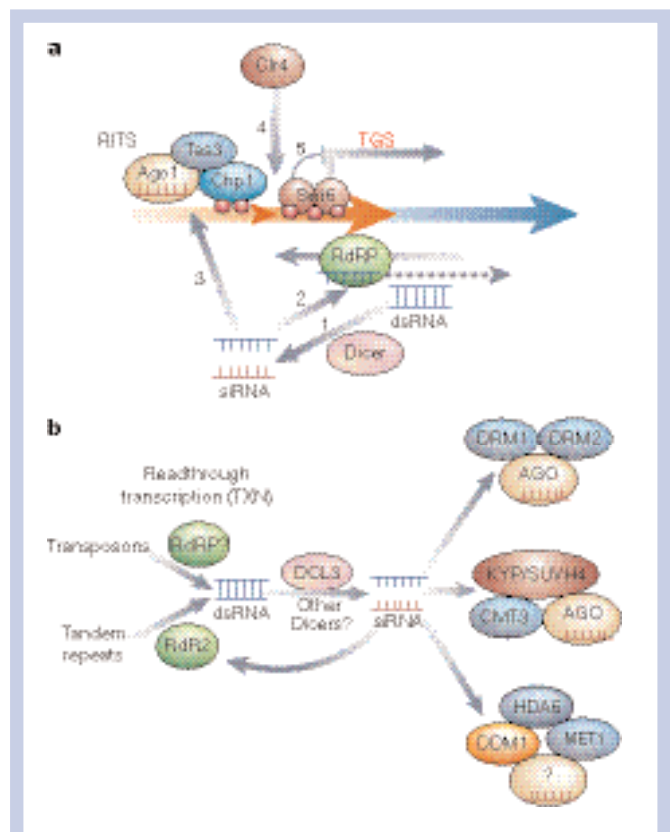


Figure 2 RNAi and heterochromatic silencing. **a**, Silencing in *S. pombe*. The reverse strand of the centromeric repeats (Box 1) is transcribed in wild-type cells, but rapidly processed into siRNA by the RNAi pathway (1). The resulting siRNAs are amplified by RdRP (2), which is recruited to the repeats. siRNAs are also found in the RITS complex (3). These siRNAs are not required for the assembly of RITS but they are required for targeting. Targeting may be mediated either directly or indirectly by means of H3mK9 (red circles), which in turn probably interacts with Chp1. The role of RITS, if any, in modification of H3K9 by Clr4 (4) is not yet understood, but this leads to recruitment of another chromodomain protein Swi6, which silences the forward strand by TGS (5). **b**, Silencing in *Arabidopsis*. Transposons and tandem repeats account for 95% of siRNA in *Arabidopsis*, at least some of which is processed by DCL3 and RdR2 (RNA-dependent RNA polymerase 2), although other enzymes may be involved. At least three groups of genes, whose products may form complexes, have been proposed to account for heterochromatic modifications that are guided by siRNA. The first includes DRM1 and/or DRM2, which encode redundant Dnmt3-like methyltransferases. These are required for *de novo* DNA methylation including non-CG methylation. The second group includes CMT3, which is required for CNG methylation and is recruited by histone H3mK9 by means of KYP and possibly other histone methyltransferases. The third group includes the Dnmt1-like methyltransferase MET1, which is required to maintain CG methylation, but may also establish methylation in the presence of siRNA. Also included in this group is the histone deacetylase HDA6, which is probably required indirectly for H3mK9. Each putative complex is thought to bind siRNA. The first two groups include at least one AGO gene, including AGO4 and AGO1. The mechanism by which siRNA accumulation depends on MET1 and DDM1 is unknown.

mouse, in that only the paternal chromosome is silenced in the extraembryonic tissues⁷⁸. Dosage compensation in *Drosophila* uses a different mechanism, but also uses noncoding roX RNA that interacts with chromodomains⁸⁰.

Finally, RNAi mutants in animals and plants frequently have stem-cell defects (see reviews in this issue by Ambros, page 350 and Baulcombe, page 356). Stem-cell divisions can be thought of as chromosomal imprints if the parental chromosome set is retained in mother cells of asymmetric stem-cell lineages — as has been indicated in some studies of mammalian cells⁸¹. It is tempting to implicate centromeric transcripts in these imprints, given that RNAi is required for

Box 3

Homology-dependent gene silencing

In plants, animals and filamentous fungi sequence-specific gene silencing occurs when multiple copies of transgenes are integrated into the genome. Endogenous genes can participate in this silencing, aspects of which are known as co-suppression⁸⁴. Silencing can be post-transcriptional (PTGS), and occur through the sequence-specific degradation of homologous RNA, or transcriptional (TGS), and occur through chromatin modifications that suppress transcription.

PTGS in plants requires *AGO1*, *RdRP* (*SDE1/SGS2*), RNA helicases (*SDE2/SGS3*) and other components required for RNAi (see review in this issue by Baulcombe, page 356). Homologues of these genes are also required for PTGS in *Neurospora* and *C. elegans*. TGS on the other hand depends on DNA methylation — primarily of the promoter regions, although methylation of coding sequences may also have a role²⁵. In the case of RIP (repeat-induced point mutations) in *Neurospora*, silencing is also associated with base transitions, which may be the result of methylation-coupled DNA damage and repair⁹⁵. Imprinting in plants can also depend on DNA repair, although the mechanism may be unrelated⁹⁶. In *Drosophila*, which has little or no DNA methylation, *alcohol dehydrogenase* (*Adh*) transgenes are silenced by Polycomb-dependent TGS, or by PTGS involving RNA⁶³.

Tandem arrays of the *white* minigene resemble heterochromatin and are subject to PEV⁶². However, aspects of TGS depend on genes required for PTGS, indicating that RNA participates in both types of silencing⁶².

Homology-dependent silencing can be mediated by DNA or RNA. DNA–DNA pairing can result in transfer of silencing and DNA methylation²⁸. Such pairing during meiosis transfers silencing and methylation in *Ascombolus*⁹⁷. In *Neurospora*, unpaired sequences are silenced meiotically, and this silencing depends on RNA-dependent RNA polymerase and not on the *dim-2* DNA methyltransferase, indicating that RNA might be an intermediate in homology sensing⁹⁸. In contrast, in mitotic cells, transgene methylation depends on *dim-2*, and not on RNA-dependent RNA polymerase or other genes required for RNA⁹⁹. It is likely, therefore, that different mechanisms are involved in each case. In plants, a role for RNA in homology-dependent DNA silencing was first demonstrated using viroid (small RNA plant pathogens) transgenes, which become methylated if RNA viroids are replicating in the cytoplasm¹⁰⁰. The discovery of siRNA provides a possible intermediate (see review in this issue by Baulcombe, page 356).

normal chromosome segregation^{33,34}. The transcribed strands are inherently asymmetric, at least in yeast³⁰ and plants (B. May *et al.*, unpublished data), providing a plausible mechanism.

Perspective

RNAi provides a mechanism for programmed, sequence-specific silencing by means of heterochromatic modifications, which is likely to be particularly important for silencing transposons and repeats. In 1951 McClintock proposed gene control by transposable elements as a potential mechanism for developmental regulation^{82,83}. Although imprinting and cellular memory may well be regulated in this way, heterochromatic silencing could prove to be important over a longer timeframe in population genetics and evolution. Heterochromatin is a dynamic component of the genome¹ and varies between natural ecotypes of most species^{83,84}. In wide crosses between diverse strains, such as those responsible for generating natural allopolyploids (individuals containing multiple chromosome sets derived from different species), siRNAs from one parent might not match polymorphic repeats from the other. If siRNA pools are largely maternal, cytoplasmic siRNA could contribute to chromosome loss and transposon activation (hybrid dysgenesis) if they fail to match paternal chromosome sequences^{85–87}. Conversely, loss of silencing in transposon-regulated genes could lead to increased gene expression, and contribute to hybrid vigour²⁵. Aspects of the RNAi-mediated heterochromatic silencing mechanism are still unclear. In particular, we know little about the polymerase responsible for generating dsRNA, or how specific sequences are targeted. However, this RNAi-dependent gene silencing mechanism will surely have widespread implications for the biological role of heterochromatin. □

doi:10.1038/nature02875

- Heitz, E. Das heterochromatin der Moose. *Jahrb. Wiss. Botanik* **69**, 762–818 (1928).
- Hennig, W. Heterochromatin. *Chromosoma* **108**, 1–9 (1999).
- Muller, H. J. Types of visible variations induced by X-rays in *Drosophila*. *J. Genetics* **22**, 299–334 (1930).
- Comfort, N. C. From controlling elements to transposons: Barbara McClintock and the Nobel Prize. *Trends Genet.* **17**, 475–478 (2001).
- Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074–1080 (2001).
- Lachner, M. & Jenuwein, T. The many faces of histone lysine methylation. *Curr. Opin. Cell Biol.* **14**, 286–298 (2002).
- Brehm, A., Tufteland, K. R., Aasland, R. & Becker, P. B. The many colours of chromodomains. *Bioessays* **26**, 133–140 (2004).
- Jainisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic

- and environmental signals. *Nature Genet.* **33** (Suppl.), 245–254 (2003).
- Tamaru, H. & Selker, E. U. A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. *Nature* **414**, 277–283 (2001).
- Lindroth, A. M. *et al.* Requirement of *CHROMOMETHYLASE3* for maintenance of CpXpG methylation. *Science* **292**, 2077–2080 (2001).
- Bartee, L., Malagnac, F. & Bender, J. *Arabidopsis cmt3* chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. *Genes Dev.* **15**, 1753–1758 (2001).
- Jackson, J. P., Lindroth, A. M., Cao, X. & Jacobsen, S. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **416**, 556–560 (2002).
- Malagnac, F., Bartee, L. & Bender, J. An *Arabidopsis* SET domain protein required for maintenance but not establishment of DNA methylation. *EMBO J.* **21**, 6842–6852 (2002).
- Tariq, M. *et al.* Erasure of CpG methylation in *Arabidopsis* alters patterns of histone H3 methylation in heterochromatin. *Proc. Natl Acad. Sci. USA* **100**, 8823–8827 (2003).
- Lippman, Z., May, B., Yordan, C., Singer, T. & Martienssen, R. Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PLoS Biol.* **1**, E67 (2003).
- Johnson, L., Cao, X. & Jacobsen, S. Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. *Curr. Biol.* **12**, 1360–1367 (2002).
- Fuks, F., Hurd, P. J., Deplus, R. & Kouzarides, T. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res.* **31**, 2305–2312 (2003).
- Fuks, F. *et al.* The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J. Biol. Chem.* **278**, 4035–4040 (2003).
- Lehnertz, B. *et al.* Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr. Biol.* **13**, 1192–1200 (2003).
- Nguyen, C. T. *et al.* Histone H3-lysine 9 methylation is associated with aberrant gene silencing in cancer cells and is rapidly reversed by 5-aza-2'-deoxycytidine. *Cancer Res.* **62**, 6456–6461 (2002).
- Jeddeloh, J. A., Stokes, T. L. & Richards, E. J. Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nature Genet.* **22**, 94–97 (1999).
- Gendrel, A. V., Lippman, Z., Yordan, C., Colot, V. & Martienssen, R. A. Dependence of heterochromatic histone H3 methylation patterns on the *Arabidopsis* gene *DDM1*. *Science* **297**, 1871–1873 (2002).
- Vongs, A., Kakutani, T., Martienssen, R. A. & Richards, E. J. *Arabidopsis thaliana* DNA methylation mutants. *Science* **260**, 1926–1928 (1993).
- Kankel, M. W. *et al.* *Arabidopsis* *MET1* cytosine methyltransferase mutants. *Genetics* **163**, 1109–1122 (2003).
- Martienssen, R. A. & Colot, V. DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science* **293**, 1070–1074 (2001).
- Lippman, Z. L. *et al.* Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**, 471–476 (2004).
- Kato, M., Miura, A., Bender, J., Jacobsen, S. E. & Kakutani, T. Role of CG and non-CG methylation in immobilization of transposons in *Arabidopsis*. *Curr. Biol.* **13**, 421–426 (2003).
- Matzke, M. *et al.* Genetic analysis of RNA-mediated transcriptional gene silencing. *Biochim. Biophys. Acta* **1677**, 129–141 (2004).
- Grewal, S. I. Transcriptional silencing in fission yeast. *J. Cell Physiol.* **184**, 311–318 (2000).
- Volpe, T. A. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833–1837 (2002).
- Reinhart, B. J. & Bartel, D. P. Small RNAs correspond to centromere heterochromatic repeats. *Science* **297**, 1831 (2002).
- Bernard, P. *et al.* Requirement of heterochromatin for cohesion at centromeres. *Science* **294**, 2539–2542 (2001).
- Volpe, T. *et al.* RNA interference is required for normal centromere function in fission yeast. *Chromosome Res.* **11**, 137–146 (2003).

34. Hall, I. M., Noma, K. & Grewal, S. I. RNA interference machinery regulates chromosome dynamics during mitosis and meiosis in fission yeast. *Proc. Natl Acad. Sci. USA* **100**, 193–198 (2003).
35. Verdel, A. *et al.* RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672–676 (2004).
36. Martienssen, R. A. Maintenance of heterochromatin by RNA interference of tandem repeats. *Nature Genet.* **35**, 213–214 (2003).
37. Jacobs, S. A. & Khorasanizadeh, S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295**, 2080–2083 (2002).
38. Schramke, V. & Allshire, R. Hairpin RNAs and retrotransposon LTRs affect RNAi and chromatin-based gene silencing. *Science* **301**, 1069–1074 (2003).
39. Hall, I. M. *et al.* Establishment and maintenance of a heterochromatin domain. *Science* **297**, 2232–2237 (2002).
40. Jia, S., Noma, K. & Grewal, S. I. RNAi-independent heterochromatin nucleation by the stress-activated ATF/CREB family proteins. *Science* **304**, 1971–1976 (2004).
41. Dernburg, A. F., Zalevsky, J., Colaiacovo, M. P. & Villeneuve, A. M. Transgene-mediated cosuppression in the *C. elegans* germline. *Genes Dev.* **14**, 1578–1583 (2000).
42. Wu-Scharf, D., Jeong, B., Zhang, C. & Cerutti, H. Transgene and transposon silencing in *Chlamydomonas reinhardtii* by a DEAH-box RNA helicase. *Science* **290**, 1159–1162 (2000).
43. Aravin, A. A. *et al.* Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr. Biol.* **11**, 1017–1027 (2001).
44. Aravin, A. A. *et al.* The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell* **5**, 337–350 (2003).
45. Djikeng, A., Shi, H., Tschudi, C. & Ullu, E. RNA interference in *Trypanosoma brucei*: cloning of small interfering RNAs provides evidence for retrotransposon-derived 24–26-nucleotide RNAs. *RNA* **7**, 1522–1530 (2001).
46. Mochizuki, K., Fine, N., Fujisawa, T. & Gorovsky, M. Analysis of a *piwi*-related gene implicates small RNAs in genome rearrangement in *Tetrahymena*. *Cell* **110**, 689 (2002).
47. Taverna, S. D., Coyne, R. S. & Allis, C. D. Methylation of histone H3 at lysine 9 targets programmed DNA elimination in *Tetrahymena*. *Cell* **110**, 701–711 (2002).
48. Xie, Z. *et al.* Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2**, E104 (2004).
49. Sijen, T. & Plasterk, R. H. Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature* **426**, 310–314 (2003).
50. Zilberman, D., Cao, X. & Jacobsen, S. E. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299**, 716–719 (2003).
51. Lynn, K. *et al.* The PINHEAD/ZWILLE gene acts pleiotropically in *Arabidopsis* development and has overlapping functions with the ARGONAUTE1 gene. *Development* **126**, 469–481 (1999).
52. Soppe, W. J. *et al.* The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol. Cell* **6**, 791–802 (2000).
53. Kinoshita, T. *et al.* One-way control of FWA imprinting in *Arabidopsis* endosperm by DNA methylation. *Science* **303**, 521–523 (2004).
54. Chan, S. W. *et al.* RNA silencing genes control *de novo* DNA methylation. *Science* **303**, 1336 (2004).
55. Cao, X. & Jacobsen, S. E. Role of the *Arabidopsis* DRM methyltransferases in *de novo* DNA methylation and gene silencing. *Curr. Biol.* **12**, 1138–1144 (2002).
56. Cao, X. *et al.* Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr. Biol.* **13**, 2212–2217 (2003).
57. Bender, J. & Fink, G. R. Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of *Arabidopsis*. *Cell* **83**, 725–734 (1995).
58. Mette, M. F., Aufsatz, W., van der Winden, J., Matzke, M. A. & Matzke, A. J. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J.* **19**, 5194–5201 (2000).
59. Melquist, S. & Bender, J. Transcription from an upstream promoter controls methylation signaling from an inverted repeat of endogenous genes in *Arabidopsis*. *Genes Dev.* **17**, 2036–2047 (2003).
60. Zilberman, D. *et al.* Role of *Arabidopsis* ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. *Curr. Biol.* **14**, 1214–1220 (2004).
61. Fagard, M., Boutet, S., Morel, J. B., Bellini, C. & Vaucheret, H. AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. *Proc. Natl Acad. Sci. USA* **97**, 11650–11654 (2000).
62. Pal-Bhadra, M. *et al.* Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery. *Science* **303**, 669–672 (2004).
63. Pal-Bhadra, M., Bhadra, U. & Birchler, J. A. RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in *Drosophila*. *Mol. Cell* **9**, 315–327 (2002).
64. Lee, Y. S. *et al.* Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**, 69–81 (2004).
65. Kelly, W. G. *et al.* X-chromosome silencing in the germline of *C. elegans*. *Development* **129**, 479–492 (2002).
66. Maison, C. *et al.* Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nature Genet.* **30**, 329–334 (2002).
67. Bernstein, E. *et al.* Dicer is essential for mouse development. *Nature Genet.* **35**, 215–217 (2003).
68. Rudert, F., Bronner, S., Garnier, J. M. & Dolle, P. Transcripts from opposite strands of gamma satellite DNA are differentially expressed during mouse development. *Mamm. Genome* **6**, 76–83 (1995).
69. Deng, W. & Lin, H. *miwi*, a murine homolog of *piwi*, encodes a cytoplasmic protein essential for spermatogenesis. *Dev. Cell* **2**, 819–830 (2002).
70. Kuramochi-Miyagawa, S. *et al.* *Mili*, a mammalian member of *piwi* family gene, is essential for spermatogenesis. *Development* **131**, 839–849 (2004).
71. Kelley, R. L. & Kuroda, M. I. Noncoding RNA genes in dosage compensation and imprinting. *Cell* **103**, 9–12 (2000).
72. Reik, W. & Walter, J. Genomic imprinting: parental influence on the genome. *Nature Rev. Genet.* **2**, 21–32 (2001).
73. Kiyosue, T. *et al.* Control of fertilization-independent endosperm development by the *MEDEA* polycomb gene in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **96**, 4186–4191 (1999).
74. Baroux, C., Spillane, C. & Grossniklaus, U. Genomic imprinting during seed development. *Adv. Genet.* **46**, 165–214 (2002).
75. Sleutels, F. & Barlow, D. P. The origins of genomic imprinting in mammals. *Adv. Genet.* **46**, 119–163 (2002).
76. Seitz, H. *et al.* Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nature Genet.* **34**, 261–262 (2003).
77. Stam, M., Bebele, C., Dorweiler, J. E. & Chandler, V. L. Differential chromatin structure within a tandem array 100 kb upstream of the maize *b1* locus is associated with paramutation. *Genes Dev.* **16**, 1906–1918 (2002).
78. Avner, P. & Heard, E. X-chromosome inactivation: counting, choice and initiation. *Nature Rev. Genet.* **2**, 59–67 (2001).
79. Heard, E. *et al.* Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell* **107**, 727–738 (2001).
80. Akhtar, A., Zink, D. & Becker, P. B. Chromodomains are protein-RNA interaction modules. *Nature* **407**, 405–409 (2000).
81. Merok, J. R., Lansita, J. A., Tunstead, J. R. & Sherley, J. L. Cosegregation of chromosomes containing immortal DNA strands in cells that cycle with asymmetric stem cell kinetics. *Cancer Res.* **62**, 6791–6795 (2002).
82. McClintock, B. Chromosome organization and genic expression. *Cold Spring Harbor Symp. Quant. Biol.* **16**, 13–47 (1951).
83. McClintock, B., Kato, Y., T. A. & Blumenschein, A. *Chromosome Constitutions of Races of Maize. Their Significance for Interpreting Relationships Among Races and Strains in the Americas (A monograph.)* (Colegio de Postgraduados, Escuela Nacional de Agricultura, Chapingo, Edo. Mexico, 1981).
84. Frantz, P., Soppe, W. & Schubert, I. Heterochromatin in interphase nuclei of *Arabidopsis thaliana*. *Chromosome Res.* **11**, 227–240 (2003).
85. Osborn, T. C. *et al.* Understanding mechanisms of novel gene expression in polyploids. *Trends Genet.* **19**, 141–147 (2003).
86. Heath, E. M. & Simmons, M. J. Genetic and molecular analysis of repression in the P-M system of hybrid dysgenesis in *Drosophila melanogaster*. *Genet. Res.* **57**, 213–226 (1991).
87. Sarot, E., Payen-Groschene, G., Bucheton, A. & Pelisson, A. Evidence for a *piwi*-dependent RNA silencing of the gypsy endogenous retrovirus by the *Drosophila melanogaster flammenco* gene. *Genetics* **166**, 1313–1321 (2004).
88. Sun, X., Le, H. D., Wahlstrom, J. M. & Karpen, G. H. Sequence analysis of a functional *Drosophila* centromere. *Genome Res.* **13**, 182–194 (2003).
89. *Arabidopsis* Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
90. Choo, K. H. Domain organization at the centromere and neocentromere. *Dev. Cell* **1**, 165–177 (2001).
91. Ananiev, E. V., Phillips, R. L. & Rines, H. W. Complex structure of knobs and centromeric regions in maize chromosomes. *Tsitol. Genet.* **34**, 11–15 (2000).
92. Nagaki, K. *et al.* Sequencing of a rice centromere uncovers active genes. *Nature Genet.* **36**, 138–145 (2004).
93. Selker, E. U. *et al.* The methylated component of the *Neurospora crassa* genome. *Nature* **422**, 893–897 (2003).
94. Jorgensen, R. A. in *RNAi: A Guide to Gene Silencing* (ed. Hannon, G. J.) Ch. 1, 5–21 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2003).
95. Kouzminova, E. & Selker, E. U. *dim-2* encodes a DNA methyltransferase responsible for all known cytosine methylation in *Neurospora*. *EMBO J.* **20**, 4309–4323 (2001).
96. Xiao, W. *et al.* Imprinting of the *MEA* Polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase. *Dev. Cell* **5**, 891–901 (2003).
97. Colot, V., Maloel, L. & Rossignol, J. L. Interchromosomal transfer of epigenetic states in *Ascomycetes*: transfer of DNA methylation is mechanistically related to homologous recombination. *Cell* **86**, 855–864 (1996).
98. Shiu, P. K. & Metzner, R. L. Meiotic silencing by unpaired DNA: properties, regulation and suppression. *Genetics* **161**, 1483–1495 (2002).
99. Freitag, M. *et al.* DNA methylation is independent of RNA interference in *Neurospora*. *Science* **304**, 1939 (2004).
100. Wassenecker, M., Heimes, S., Riedel, L. & Sanger, H. L. RNA-directed *de novo* methylation of genomic sequences in plants. *Cell* **76**, 567–576 (1994).

Acknowledgements We thank our collaborators V. Colot and A.-V. Gendrel, and members of our laboratory for discussions. We thank J. Birchler and S. Elgin for permission to reproduce Fig. 1, and A. Frary for comments on the manuscript. Z.L. is the recipient of an Arnold and Mabel Beckman Graduate Student Fellowship of the Watson School of Biological Sciences. R.M. is supported by an NSF Plant Genome Program grant and an NIH grant.

Competing interests statement The authors declare that they have no competing financial interests.

Unlocking the potential of the human genome with RNA interference

Gregory J. Hannon¹ & John J. Rossi²

¹Watson School of Biological Sciences, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA (email: hannon@cshl.edu) ²Division of Molecular Biology, Beckman Research Institute of the City of Hope, Graduate School of Biological Sciences, Duarte, California 91010, USA (email: J.Rossi@coh.org)

The discovery of RNA interference (RNAi) may well be one of the transforming events in biology in the past decade. RNAi can result in gene silencing or even in the expulsion of sequences from the genome. Harnessed as an experimental tool, RNAi has revolutionized approaches to decoding gene function. It also has the potential to be exploited therapeutically, and clinical trials to test this possibility are already being planned.

The formal description of RNAi as a biological response to double-stranded RNA (dsRNA) came about following experiments with dsRNA in the nematode *Caenorhabditis elegans*^{1,2}. Injecting dsRNAs into the worm was found to silence genes whose sequences were complementary to those of the introduced dsRNAs³. It is now clear that an RNAi pathway is present in many, if not most, eukaryotes⁴. dsRNAs are processed into short interfering RNAs (siRNAs), about 22 nucleotides in length, by the RNase enzyme Dicer. These siRNAs are then incorporated into a silencing complex called RISC (RNA-induced silencing complex), which identifies and silences complementary messenger RNAs.

Before RNAi could be harnessed as an experimental tool for silencing specific genes in mammalian systems, a considerable hurdle had to be overcome. The problem lay in making exogenous dsRNA trigger silencing in a gene-specific manner without invoking nonspecific responses to foreign dsRNAs that are part of the cell's antiviral mechanism⁵. Work over the past two years has allowed investigators to meet this challenge, and RNAi has now been adopted as a standard methodology for silencing the expression of specific genes in mammalian cells. Here, we chronicle the development of RNAi as a genetic tool in mammals, focusing on recent advances and providing practical advice for its experimental application. We also make predictions about the potential future of RNAi as a potent and specific therapeutic tool that may escape some of the limitations of conventional medicinal chemistry.

Breaking the barrier to RNAi in mammals

For more than 30 years, it has been known that exposure of mammalian cells to long dsRNAs induces innate immune pathways, including interferon-regulated responses that serve as antiviral mechanisms. The enzyme dsRNA-dependent protein kinase (PKR) is activated on binding to dsRNA and localized, but sequence-independent destruction of RNAs and a generalized repression of protein synthesis results⁶. The existence of these innate immune pathways seemed incompatible with the use of dsRNA for silencing a particular target gene. However, evidence of an RNAi pathway in mammals came from the observation that key biochemical components of the RNAi pathway are conserved^{6,7}. It was also shown that long dsRNAs can trigger gene-specific responses when they are introduced into mammalian embryos and embryonic cell lines in which nonspecific antiviral responses to dsRNAs are not prevalent⁴. This raised the problem of how to shift the response of a mammalian cell to foreign dsRNA from the non-

specific sequence-independent defence pathways to the sequence-specific RNAi pathway. Attempts to meet this challenge have resulted in RNAi being established as a genetic tool in mammalian cells and animals.

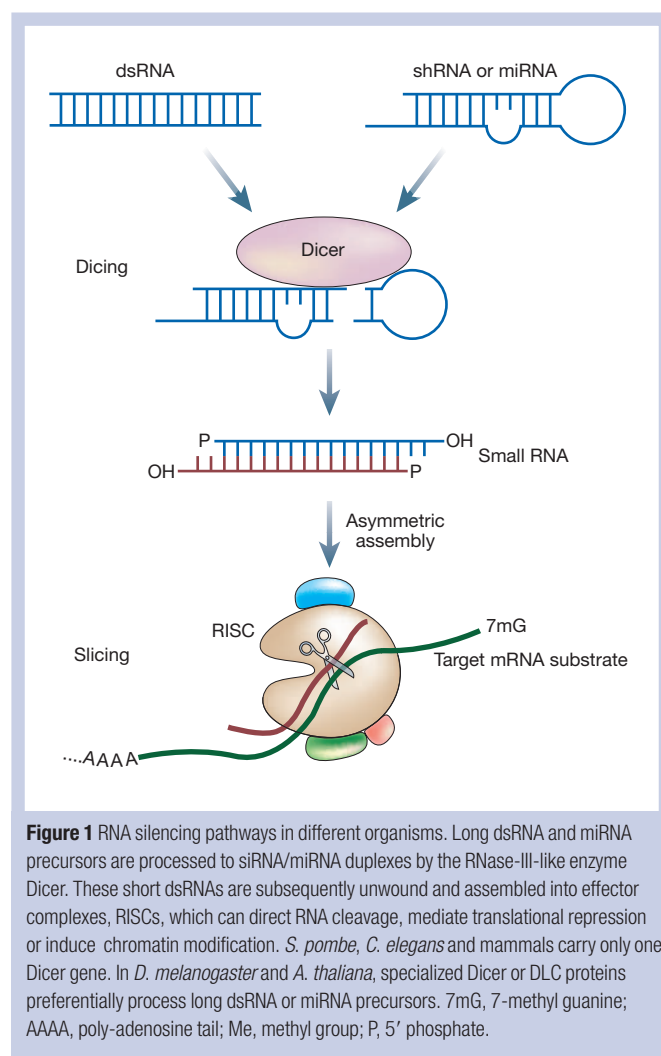
Using siRNAs for RNAi

A biochemical understanding of the RNAi pathway (Fig. 1; see review in this issue by Meister and Tuschl, page 343) was crucial to realizing that dsRNAs shorter than 30 base pairs (bp) could be used to trigger an RNAi response in mammals. Tuschl and colleagues showed that transfection of mammalian cells with short RNAs could induce the sequence-specific RNAi pathway, and so overcame the barrier to the use of RNAi as a genetic tool in mammals⁸. The impetus to use siRNAs and other small RNAs in mammalian cells also came from the long-standing view that PKR activation and similar responses were not effectively triggered by short dsRNAs. Following the initial reports, it took a remarkably short period of time for siRNAs to be adopted as a standard component of the molecular biology toolkit.

siRNAs can be introduced into mammalian cells using a variety of standard transfection methods. The strength and duration of the silencing response is determined by several factors: on a population basis, the silencing response is affected mainly by the overall efficiency of transfection, which can be addressed by optimizing conditions. In each cell, silencing depends on the amount of siRNA that is delivered and on the potential of each siRNA to suppress its target, or its potency. Even a relatively impotent siRNA can silence its target provided that sufficient quantities of the siRNA are delivered. However, essentially 'forcing' the system by delivering large amounts of reagent is likely to lead to numerous undesired effects (see section 'Intrinsic limits on the specificity of RNAi').

Using shRNAs for RNAi

The discovery of the endogenous triggers of the RNAi pathway in the form of small temporal RNAs — now termed microRNAs (miRNAs)^{9–11} — suggested that RNAi might be triggered in mammalian cells by synthetic genes that express mimics of endogenous triggers. Several laboratories simultaneously used related approaches to test this idea. These involved expressing mimics of miRNAs in the form of short hairpin RNAs (shRNAs) from RNA polymerase II or III promoters¹². The shRNAs themselves varied in size and design, with stems ranging from 19 to 29 nucleotides in length, and with various degrees of structural similarity to natural miRNAs. All these approaches were



effective to varying degrees, and at present, no real consensus has developed on the most effective way to present synthetic miRNAs to the RNAi pathway.

Because these triggers are encoded by DNA vectors, they can be delivered to cells in any of the innumerable ways that have been devised for delivery of DNA constructs that allow ectopic mRNA expression. These include standard transient transfection, stable transfection and delivery using viruses ranging from retroviruses to adenoviruses. Expression can also be driven by either constitutive or inducible promoter systems¹².

Recent studies strongly indicate that each shRNA expression construct gives rise to a single siRNA (D. Siolas, G.J.H. and M. Cleary, unpublished work). Knowing precisely how this processing occurs for each shRNA cassette design has permitted the application of siRNA design criteria (see 'Features of effective siRNAs and shRNAs' below) to the design of effective shRNAs. The use of siRNAs and shRNAs are complementary approaches in the application of RNAi as a genetic tool in mammals, and the best approach depends on the type of study being performed.

Features of effective siRNAs and shRNAs

Observations of widely varying efficacy of individual siRNAs motivated a search for rules that might specify more effective siRNAs. Several groups took a 'black box' approach, which involved assaying large numbers of siRNAs, sorting them into classes depending on their potency, and then looking for characteristics that distinguished effective siRNAs from ineffective ones^{13–15}. Some common rules have begun to emerge from these studies. siRNAs in which the helix at the

5' end of the antisense strand has a lower stability than the 3' end of the siRNA are generally more effective silencers than those with the opposite arrangement. A biochemical basis for the thermodynamic arrangement of effective siRNAs was provided by biochemical studies of the mRNA-cleavage complex, RISC (Fig. 1), in *Drosophila* embryo extracts. These studies showed unequal incorporation of the two strands of the siRNA into RISC¹⁴. Moreover, strand biases could be manipulated by altering the thermodynamic stability of the terminal nucleotides in a way that precisely matched the rules derived from empirical studies. Finally, an examination of miRNAs, most of which produce RISC-like complexes containing only one strand of the precursor (see review in this issue by Meister and Tuschl, page 343), showed the same pattern of thermodynamic asymmetry as that shown by effective siRNAs^{13,14,16,17}. The rules for specifying effective siRNAs uncovered by these studies imply that the effectiveness of an RNAi response triggered by an siRNA is strongly dependent on siRNA structure and determined at the step of RISC assembly, during which the asymmetry in the dsRNA must be sensed and a single strand chosen for productive incorporation into the enzyme. Once the active RISC is formed, it is relatively insensitive to the placement or structure of the target site within the mRNA.

Intrinsic limits on the specificity of RNAi

Although RNAi silences gene expression in a sequence-specific manner, several recent studies have suggested that the specificity of silencing is not absolute. Off-target effects in mammals can come from several different sources. As discussed previously, transfection of cells with dsRNAs can activate innate immune pathways. PKR activation was thought to depend on the length of the dsRNA, with a minimal cut-off for PKR activation being roughly 30 bp of duplex. However, recent reports have suggested that both siRNAs and shRNAs can — under some circumstances and in certain cell types — activate a PKR response^{18–20}. Furthermore, siRNAs transcribed *in vitro* using bacteriophage polymerases can be potent activators of an interferon response if the initiating triphosphate is not completely removed from the transcripts²¹. Further studies are required to investigate the frequency with which RNAi triggers provoke these antiviral response pathways, and the sequence or structural characteristics that might lead an siRNA or an shRNA to trigger such a response.

miRNAs recognize and regulate their targets despite a lack of perfect complementarity. This raises the possibility that siRNAs might also not require contiguous base pairing to suppress their targets effectively; several microarray studies suggested that siRNAs can provoke sequence-dependent, off-target effects and that these can be elicited by 14 base pairings, or possibly even fewer, between the siRNA and its target²². Notably, analysis of such interactions suggested that base pairing at the 5' end of the siRNA contributed disproportionately to targeting, a conclusion also reached by analysis of interactions between miRNAs and their validated targets^{23–26}. Although such information can aid the design of more specific siRNAs, we do not have sufficient understanding of target recognition by RISC to say with certainty that we can eliminate off-target effects. In fact, the intrinsic specificity of the RNAi pathway may be sufficiently low to prevent the design of a completely specific siRNA in mammals. Fewer studies have been carried out with shRNA expression cassettes, but similar caveats undoubtedly apply.

Off-target effects can also occur at the level of protein synthesis. miRNAs in animals often regulate protein expression without having corresponding effects on mRNA levels. Several studies have indicated that siRNAs also do this, provided that mRNA cleavage is blocked by altering the geometry of the target–substrate interaction^{24–26}. However, suppression of a reporter gene in this manner was only effective when several siRNA binding sites were present. Although these findings might provide some degree of comfort to those using siRNAs experimentally, a recent study suggested that changes in the expression levels of a large number of proteins occurred in cells treated

with siRNAs. However, it should be noted that the siRNAs in question were relatively impotent²⁷. Therefore, we must view unwanted changes in protein expression levels as the 'monster under the bed' for RNAi-based studies of gene function. Ultimately, caveats in the specificity of the RNAi response make it essential to follow relatively simple guidelines for good experimental practice. These are outlined in Box 1.

RNAi as a solution for mammalian genetics

One of the first choices to make in any RNAi-based genetic experiment is whether to trigger suppression through the use of siRNAs or shRNAs. The advantages of using siRNAs are relative ease of availability and high efficiency of delivery. In addition, pre-validated siRNAs are becoming increasingly available as commercial suppliers and the scientific community acquire more experience. Overall, siRNA delivery is likely to result in the highest intracellular concentration of the gene silencer. But limitations to the use of siRNAs are that their effects are transient and restricted by the rate of cell division: mammalian cells do not have mechanisms to amplify and propagate RNAi (unlike *C. elegans* and plants). In addition, some cell types are notoriously difficult to trans-

fect, and the procedure of transfection itself can alter the physiology of the cell. However, despite these drawbacks, transfection of siRNAs is probably the fastest and easiest method currently available for producing a knockdown of gene expression in cell culture by means of RNAi.

With shRNAs, the up-front investment is greater. First, DNA oligonucleotides must be cloned and sequenced so that a construct can be produced. Second, the shRNA must be designed effectively, and consensus on the most effective design, with respect to either the structure of the shRNA itself or the structure of the expression vector, is only just beginning to emerge. However, shRNAs are capable of producing sustained repression, and allow for delivery by conventional transfection or by several advanced viral vectors that also permit stable integration into the genome. In addition, shRNA expression vectors can be propagated indefinitely. As with siRNAs, design algorithms can be applied to shRNAs to maximize the probability of success in a suppression experiment. However, the application of such algorithms requires a detailed understanding of the vector system being used.

Both siRNAs and shRNAs have been used for studies of gene function *in vivo*, primarily in mice. Both types of trigger can be

Box 1

Rules of the road for effective RNAi experiments

Given the significant concerns about the specificity of RNAi-mediated repression, how can investigators maximize confidence in the results of studies that use these tools? It is important to note that no approach used to inactivate gene function is free from potential problems. Even conventional gene knockouts are known to be subject to compensation during development. Thus, the enthusiasm for the use of RNAi as a genetic tool should be tempered by a recognition of the potential problems and good practices should be applied to avoid misinterpreting results. Four guidelines for good practice in RNAi experiments in mammals are presented below.

1. Get the right strand into RISC by using good design

RNAi-based experiments will be more informative and go more smoothly if effective and highly specific RNA triggers are used. Many algorithms now exist for choosing effective sequences. In addition, homology to other sequences in the genome should be minimized, with particular attention to the 5' end of the antisense strand. Use of design algorithms based on thermodynamic criteria can aid biased incorporation of the antisense strand of the siRNA into RISC. Several public websites provide support for such designs (see for example <http://web.mit.edu/mmcmmanus/www/home1.2files/siRNAs.htm>; <http://hydra1.wistar.upenn.edu/Projects/siRNA/siRNAindex.htm>; <http://www.cshl.edu/public/SCIENCE/hannon.html>).

2. Several alleles are better than one

Several siRNAs or shRNAs should give the same phenotypic outcome, as it is extremely unlikely that different triggers will have the same off-target effects²². It is critical to correlate this phenotypic outcome with the effectiveness of suppression. Only effective siRNAs against a given target, but not ineffective siRNAs, should yield similar phenotypes. Importantly, discrimination between effective and ineffective siRNAs can only be accomplished by examining target protein levels. There are numerous anecdotal reports of siRNAs effectively suppressing protein production without changing mRNA levels. In addition, siRNAs or shRNAs that do not affect the target protein should be used as negative controls. Arguably, one could use a 'scrambled' siRNA or shRNA for this purpose. However, such scrambled siRNAs may not have any biological activity, and it is undoubtedly better to use an RNA that is known to enter the RNAi pathway effectively. For example, an RNA targeting luciferase, green fluorescent protein or another marker gene (that is validated against its target) would be expected to enter RISC but would not be

expected to affect the expression of proteins in a mammalian cell. Other possible controls include an RNA with flipped asymmetry. This could be achieved by creating an siRNA with a more stable helix at the 5' end of the antisense strand.

3. Work at the lowest possible concentrations

RISC is a conventional enzyme, and working at high enzyme to substrate ratios is likely to affect its specificity. Therefore, it is important to identify RNAi triggers that work at very low effective concentrations. With siRNAs, this can be achieved by titrating siRNA concentrations and by correlating their effects on phenotypic outcome with both the concentration of the siRNA used and with the degree of suppression obtained. For example, if the siRNA shows maximal suppression at 5 nM but the phenotype is not observed until the concentration reaches 100 nM, off-target effects must be suspected. In fact, some siRNAs in HeLa cells have shown IC₅₀ values (the amount of siRNA required to suppress the target to 50% of its original level) of as little as 500 pM. Similarly, titration of shRNA-expression vectors should also be performed.

4. Rescue to the rescue

Ultimately, the best experiments demonstrate that expression of a version of the targeted gene that cannot be recognized by the siRNA reverts the phenotype. This can be achieved in several ways. First, mutations can be introduced into a cDNA encoding the targeted gene that destroy complementarity with the siRNA or shRNA while maintaining the wild-type protein sequence. Alternatively, the phenotype can be validated by using siRNAs or shRNAs that target untranslated regions, and then by rescuing the phenotype with an expression construct containing only the coding sequence. Although rescue experiments provide the ultimate test of the specificity of a given effect, these can be problematic. For example, it may be difficult to achieve appropriate expression levels of a particular protein. Overexpression could cause artefactual effects (for example, a pathway could be rescued by bypassing its requirements, rather than truly reverting a specific effect).

Ultimately, as our understanding of the RNAi pathway deepens, we will be able to predict with good accuracy all the on- and off-target effects of siRNAs. This will allow not only the generation of RNAi triggers with maximal specificity, but also the design of triggers that are directed against the most likely off-target genes for each siRNA or shRNA.

delivered transiently. The first demonstration of RNAi-mediated repression in an adult animal showed effective repression of a luciferase reporter gene following hydrodynamic transfection of siRNAs or shRNA expression plasmids into mouse liver²⁸. Subsequent studies have delivered siRNAs or shRNAs by various methods, including lipid-based delivery and naked RNA or DNA injection^{29–32}.

Long-term gene silencing has been demonstrated *in vivo* using both genetic mosaics and germline modification. For example, the growth of a tumour cell line in a xenograft model can be attenuated by engineering that cell line with an shRNA cassette that targets the activated *ras* oncogene before the tumour cells are subcutaneously injected into the host animal³³. Genetically mosaic animals have been created by engineering stem cells with shRNA vectors and then by using those stem cells to repopulate an organ system^{34,35}. Strains of mice have been engineered to heritably suppress a targeted gene based on inheritance of a dominantly acting shRNA expression cassette^{36–39}. Several approaches have been used to create such strains, including standard nuclear injection, creation of chimaeras with engineered embryonic stem cells, and transgenesis mediated by subzonal injection of fertilized eggs with recombinant lentiviruses. Ultimately these developments will rapidly lead to the creation of animals with inducible, tissue-specific silencing of almost any gene. RNAi is therefore likely to complement existing large-scale efforts to functionally map the mouse genome by chemical or insertional mutagenesis. RNAi is certainly complementary to such approaches, because each approach can generate different types of allele. However, unlike mutational approaches RNAi has the potential to be extended beyond mice into animals where recombinant organisms cannot be generated using embryonic stem cells.

RNAi as a tool for genome-wide studies

The success in using RNAi for analysing single genes has led inevitably to efforts to apply this approach on a large scale for forward genetics (whereby mutant genes are isolated from organisms showing abnormal physical and behavioural characteristics). Indeed, given the recent completion of the human, mouse and rat genomes, RNAi provides a ready mechanism by which this enormous wealth of sequence information can be translated into functional definitions for every gene.

Genome-wide libraries of siRNAs can be constructed in fundamentally different ways, including chemical synthesis or enzymatic digestion of long dsRNAs. An example of progress towards this goal can be found in a small-scale effort⁴⁰ to target genes in the phosphatidylinositol 3-OH kinase (PI(3)K) pathway in which a mini-library (148 siRNAs) was searched for genes that affected phosphorylation of Akt, a downstream substrate for PI(3)K.

Alternatively, libraries can be produced by constructing shRNA expression vectors that target each gene. As with siRNAs, proof of principle came initially from small-scale efforts. Using a library directed against the family of de-ubiquinating enzymes, the tumour suppressor CYLD (encoded by the familial cylindromatosis susceptibility gene) was identified as a suppressor of NF- κ B activity⁴¹. This resulted to proposals for treating cylindromatosis with existing drugs and provided powerful confirmation that unbiased, genetic approaches can lead not only to new insights in biology but also to practical advances in the treatment of disease. Two groups have recently reported the production of arrayed libraries from chemically synthesized oligonucleotides that cover about 10,000 different human genes each^{42,43}. Another group has generated a library of polymerase chain reaction (PCR) products that encode shRNAs⁴⁴, and several groups have reported methods for constructing random shRNA libraries based on manipulation of complementary DNA or genomic DNA^{45–47}. Each approach has specific advantages. Random libraries are relatively inexpensive to produce and can cover an individual gene with many different shRNAs. However, they suffer from a lack of normal representation of all genes. Libraries produced from chemically synthesized oligonucleotides are expensive. However, they permit the use of powerful informatic tools to aid

shRNA design and allow flexibility in optimizing the structure of the shRNA for entry into the RNAi pathway. In addition, synthetic libraries can be used either as mixtures or as individual arrays, in a similar way to siRNA libraries.

Large-scale screening using siRNA libraries must be carried out by individual transfection and phenotypic characterization of target cells (Fig. 2). As such, siRNA libraries can be applied to the wide range of screening methods that are being developed by the pharmaceutical industry in the form of cell-based assays for drug development. These include fluorescent reporter screens, assays for various activities in cell lysates and screening by means of automated microscopy. Alternatively, RNAi triggers can be printed on microarrays and tested for their effects following transfection *in situ*^{48,49}. Arrayed libraries of shRNAs can be used in a similar fashion: this was demonstrated by applying an arrayed library to a search for genes that affect proteasome function⁴³. shRNA libraries can also be assayed, following their integration into the genomes of target cells, in pools using protocols that filter populations based on phenotypic criteria, such as a growth selection (Fig. 2). Such a test of one shRNA library yielded new links between several genes and the p53 tumour-suppressor pathway⁴². A conceptually more complicated application of pooled screens involves using molecular 'barcodes' to track how individual shRNAs behave as members of complex populations (Fig. 2).

Clearly, large-scale library efforts will evolve with our advancing understanding of the RNAi pathway. As the quality of resources improves, there will be opportunities to progress from relatively straightforward screening protocols in cultured cells to more complex genetics in whole animals.

RNAi in drug discovery and disease therapy

RNAi has begun to produce a paradigm shift in the process of drug discovery. With the large-scale screening approaches described above, RNAi can winnow lists of potential drug targets so that efforts can be focused on the most promising candidates. Moreover, since the first description of RNAi in mammalian cells, there have been numerous studies aimed towards using RNAi to treat disease. The strong appeal of RNAi in therapeutics is the potency and specificity with which gene expression can be inhibited. The possible targets for various diseases range from oncogenes to growth factors and single nucleotide polymorphisms (SNP). There is also potential for using RNAi for the treatment of viral diseases such as those caused by the hepatitis C virus (HCV) and the human immunodeficiency virus (HIV). Despite the excitement and some early proofs of principle in the literature, there are important issues and concerns about the therapeutic application of this technology, including difficulties with delivery and uncertainty about potential toxicity. However, proposals for clinical trials using either synthetic siRNAs or viral-vector-delivered shRNAs have been put forward — although none has yet been approved.

RNAi as a treatment for HIV

The development and use of double and triple drug combinations for the treatment of HIV infection has led to dramatic improvements in the lives of HIV-infected individuals. But despite the apparent successes of the new anti-retroviral drugs there are the emerging problems of drug-resistant viral variants and toxicities of the combination drugs now in use. Therefore, there is still great interest in exploring new antiviral therapeutic approaches. HIV was the first infectious agent targeted by RNAi, perhaps because the lifecycle and pattern of gene expression of HIV is well understood. Synthetic siRNAs and expressed shRNAs have been used to target several early and late HIV-encoded RNAs in cell lines and in primary haematopoietic cells including the TAR element⁵⁰, tat^{51–53}, rev^{51,52}, gag^{54,55}, env⁵⁵, vif⁶⁰, nef⁵⁰, and reverse transcriptase⁵³.

Despite the success of RNAi-mediated inhibition of HIV-encoded RNAs in cell culture, targeting the virus directly represents a substantial challenge for clinical applications because the high viral mutation rate

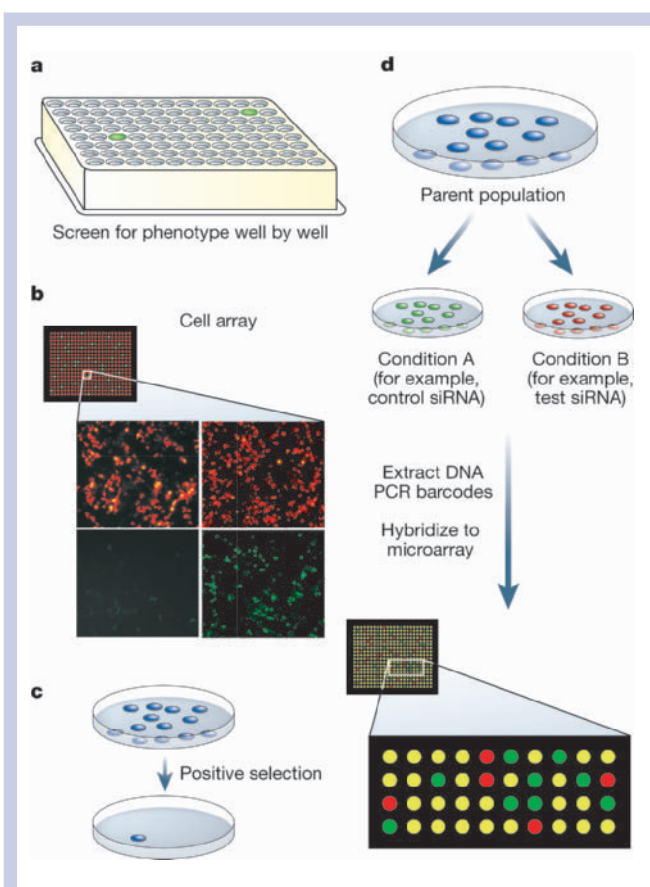


Figure 2 Genome-wide screens using RNAi. **a**, Standard methodologies can be used to screen siRNAs or shRNAs individually in 96-well plates using morphological readouts, reporters or biochemical assays. **b**, Similar approaches can be taken using siRNAs or shRNAs that have been printed on microarrays for reverse transfection⁴⁹. Reverse transfection involves the deposition of lipid–nucleic-acid complexes on a solid surface, often a glass microarray slide. Cells plated on top of the slide take up the encapsulated DNA or RNA, and this can direct mRNA expression or gene silencing. The enlarged image shows cell populations, which would be observed within individual spots of the array, expressing fluorescent proteins, red and green. In the left panels, expression of GFP has been ablated by co-deposition of a GFP siRNA. **c**, Complex (mixed) populations of shRNA-expression vectors can be filtered through positive selections in cultured cells. Selection can be for drug resistance, cytokines, genetic alterations, or — in the case of fluorescence activated cell sorter (FACS)-based selection — for cells that activate a particular marker. **d**, Complex populations can be monitored using molecular barcodes that track individual shRNA-expressing cells and their responses to various stimuli. The use of molecular barcodes combines the advantages of well-by-well screens with the advantages of carrying out pooled selections, thereby allowing the identification of phenotypes in complex populations, which do not necessarily confer a growth advantage (for example, a synthetic lethal phenotype). Each vector is tagged with a unique sequence, which can comprise the shRNA itself or a separate random or selected barcode. The frequency (representation) of each vector in a mixed population can be measured by hybridizing barcodes to an oligonucleotide microarray. If the population is subjected to selective pressure, the representation of individual shRNA constructs is expected to change as a result. This change can be detected by comparing hybridization signals for the starting population with those of the population exposed to selection. The relative signal of shRNAs that increase resistance to the selection will increase, whereas the relative signal of those that sensitize to the selection will decrease.

will lead to mutants that can escape being targeted⁵⁶. Therefore RNAi-mediated downregulation of the cellular cofactors required for HIV infection is an attractive alternative or complementary approach. Cellular cofactors such as NF- κ B⁵³, the HIV receptor CD4⁵⁴, and the co-receptors CXCR4 and CCR5⁵⁷ have been successfully downregulated

by RNAi, resulting in the inhibition of HIV replication in numerous human cell lines and in primary cells including T lymphocytes and haematopoietic-stem-cell-derived macrophages^{50–52,54,57–60}. Although targeting NF- κ B is not appropriate as a therapy owing to the important role NF- κ B has in the cell (for example, it mediates interferon-induced gene expression) the macrophage-tropic CCR5 co-receptor holds particular promise as a target. This receptor is not essential for normal immune function, and individuals homozygous for a 32-bp deletion in this gene are resistant to HIV infection, whereas individuals who are heterozygous for this deletion show delayed progression to AIDS^{61,62}. Qin *et al.*⁶³ used a lentiviral vector to transduce an anti-CCR5 shRNA in human lymphocytes. Downregulation of CCR5 resulted in a modest, but nevertheless significant three- to sevenfold reduction in viral infectivity relative to controls. Despite this downregulation, the CCR5-shRNA-treated cells were still susceptible to infection by the T-tropic virus that uses CXCR4. However, because CXCR4 is essential for the normal function of haematopoietic stem cells⁶⁴, targeting this receptor is not a good choice for an anti-HIV therapy, nor is targeting the essential CD4 receptor. So there are drawbacks in targeting cellular HIV cofactors because non-infected cells will inevitably be targeted as well, leading to toxicities that are similar to those observed with the current anti-retroviral drugs. Viral targets will need to be included in any successful strategy using RNAi. These targets should be sequences that are highly conserved throughout the various clades to ensure efficacy against all viral strains.

The delivery of siRNAs or shRNAs to HIV-infected cells is also a challenge. The target cells are primarily T lymphocytes, monocytes and macrophages. As synthetic siRNAs do not persist for long periods in cells, they would have to be delivered repeatedly for years to effectively treat the infection. Systemic delivery of siRNAs to T lymphocytes is probably not feasible owing to the immense number of these cells. Using viral vectors to deliver anti-HIV-encoding shRNA genes is also problematic, and systemic delivery is not yet practicable because the immunogenicity of the vectors themselves precludes performing multiple injections. Therefore the preferred method is to isolate T cells from patients; these T cells are then transduced, expanded and re-infused into the same patients. In a continuing clinical trial, T lymphocytes from HIV-infected individuals are transduced *ex vivo* with a lentiviral vector that encodes an anti-HIV antisense RNA. The transduced cells are subsequently expanded and reinfused into patients^{65,66}. This type of therapeutic approach would also be applicable to vectors harbouring genes that encode siRNAs. A different approach is to transduce isolated haematopoietic progenitor or stem cells with vectors harbouring the therapeutic genes. These cells give rise to all the haematopoietic cells capable of being infected by the virus. Haematopoietic stem cells are mobilized from the patient and transduced *ex vivo* before reinfusion (Fig. 3). Two clinical trials in which retroviral vectors expressing ribozymes were transduced into haematopoietic stem cells have demonstrated the feasibility of this approach^{67,68}. Because RNAi is more potent than ribozyme or antisense approaches, movement of this technology to a human clinical trial for HIV treatment is expected to take place in the next year or two.

RNAi to treat viral hepatitis

Hepatitis induced by the hepatitis B virus (HBV) and by HCV is a major health problem. At present hundreds of millions of individuals are infected worldwide. There is an effective vaccine against HBV, but this treatment is only useful for the prevention of viral infection and there is no vaccine for HCV. Therefore, hepatitis caused by these two viruses has been an important target for potential RNAi therapy. The first demonstration of RNAi efficacy against a virus *in vivo* involved hydrodynamic co-delivery of an HBV replicon and an expression unit encoding an anti-HBV shRNA in mice⁶⁹. This study demonstrated that a significant knockdown (99%) of the HBV core antigen in liver hepatocytes could be achieved by the shRNA, providing an important proof of principle for future antiviral applications of RNAi in the liver.

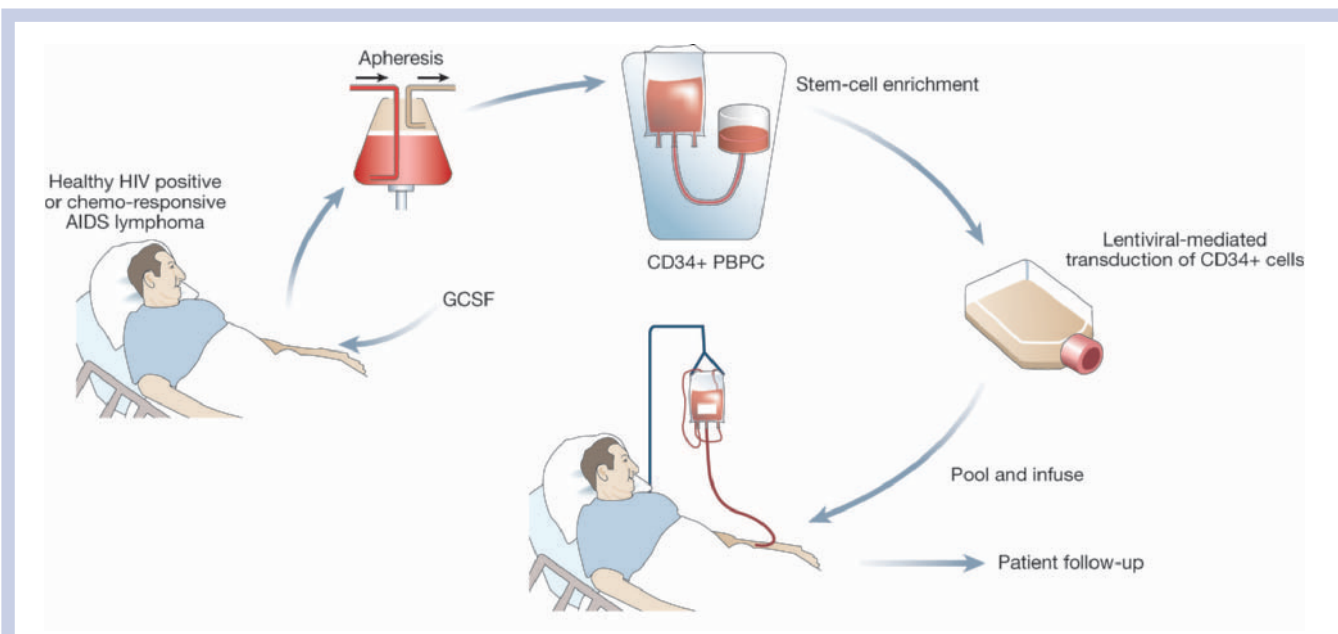


Figure 3 Proposed scheme for the treatment of HIV patients using lentiviral vectors to transduce anti-HIV shRNA genes into the patient's haematopoietic stem cells. Patients are given several injections of granulocyte colony stimulation factor (G-CSF), which mobilizes haematopoietic stem cells into the patients' peripheral circulation. Haematopoietic stem cells expressing the CD34 antigen are collected by affinity columns (apheresis) and transduced with a lentiviral vector harbouring the anti-HIV shRNA genes. The cells are then re-infused into patients. Depending on the population, the patient will have been pretreated with no, or with one, or with more than one marrow-chemoablative agent. Following stem-cell engraftment, patients are monitored for a period of several years for HIV loads, CD4+ T-lymphocytes and shRNA gene expression. This overall scheme follows that described by Michienzi *et al.*⁶⁸. PBPC, peripheral blood progenitor cells.

More advanced studies have been carried out for RNAi therapies against HCV, a virus that now infects an estimated 3% of the world's population. HCV is a major cause of chronic liver disease, which can lead to liver cirrhosis and hepatocellular carcinoma. The HCV genome is a positive-strand RNA molecule with a single open reading frame encoding a polypeptide that is processed post-translationally to produce at least ten proteins. The only therapy currently available is a combination of interferon and ribavirin, but response to this therapy is often poor, particularly with certain HCV subtypes.

Subgenomic and full-length HCV replicons that replicate and express HCV proteins in stably transfected human hepatoma-derived Huh-7 cells have been used to study the effects of various antiviral drugs^{70–73}. Several groups have now tested the efficacy of siRNA mediated inhibition of replicon function using these systems^{74–76}. siRNAs targeting the internal ribosome entry site (IRES) or mRNAs encoding the viral non-structural proteins NS3 and NS5B inhibited HCV replicon function in cell culture⁷⁵. Furthermore, anti-HCV siRNAs depleted Huh-7 cells of persistently replicating HCV replicons⁷⁴. McCaffrey *et al.* performed hydrodynamic tail-vein injections of siRNAs or anti-HCV shRNAs to direct efficient cleavage of HCV sequences in an HCV-luciferase fusion construct in mouse hepatocytes *in vivo*²⁸.

In another *in vivo* study, siRNAs were used to treat fulminant hepatitis induced by an agonistic Fas-specific antibody in mice. Anti-Fas siRNAs were hydrodynamically injected into the antibody-treated mice: 82% of the mice survived for 10 days of observation whereas all control mice died within 3 days⁷⁷. Importantly, mice already suffering from auto-immune hepatitis also improved after the Fas siRNA treatment. So it may be feasible to use siRNAs to ameliorate the severity of certain diseases by targeting the inflammatory response pathways rather than the infectious agent.

As with HIV therapeutics, delivery of the siRNAs or shRNA vectors is the main challenge for successful treatment of HCV. The method of delivery used in several *in vivo* studies — hydrodynamic intravenous injection — is not feasible for the treatment of human hepatitis. In mice, genetic material can be introduced into hepatocytes using catheters or even localized hydrodynamic procedures⁷⁸, but it is yet

to be determined whether such procedures can be used to deliver siRNAs in larger mammals.

RNAi and cancer

Many studies have used siRNAs as an experimental tool to dissect the cellular pathways that lead to uncontrolled cell proliferation and to cancer. Moreover, RNAi has been proposed as a potential treatment for cancer^{79–81}. Although no clinical trials are yet underway, a precedent might be set by ongoing clinical trials that use antisense reagents. The first systemically delivered antisense oligonucleotide for the treatment of cancer, Genasense (Genta, Inc.), which targets the anti-apoptotic gene *BCL2*, has shown promise in clinical trials for metastatic melanoma when used in combination with conventional chemotherapeutics⁸². However, its use as a US Food and Drug Administration (FDA)-approved drug has recently been put on hold. The potential for using RNAi to treat metastatic cancers will of course depend on finding good cellular targets.

Highly efficient mechanisms for the delivery of siRNA to the relevant cells will also be particularly important for successful treatment of metastatic cancer. Several groups have developed backbone modifications to synthetic siRNAs that provide them with resistance to serum nucleases and should therefore increase the half-life of circulating siRNAs in animal models^{83,84}. However, enhancing siRNA stability is not enough unless the siRNAs can penetrate cells and tissue *in vivo* in concentrations sufficient to be therapeutically functional. As siRNAs are double-stranded molecules, delivery and cellular uptake is more of a challenge than for single-stranded antisense agents, which bind to serum proteins and are taken up by cells and tissues *in vivo*⁸⁵. There are a few reports of functional RNAi being obtained by systemic delivery of liposome-encapsulated siRNAs, but the use of cationic or anionic lipids for *in vivo* delivery of antisense agents has never reached a clinical trial. Therefore, we still need to understand better which backbone modifications might be useful for enhancing cellular and tissue uptake of naked RNAs, or we need to develop alternative carriers for systemic delivery of siRNAs — a feat that will be essential in treating metastatic cancers.

Using RNAi to target genes expressing oncogenic fusion proteins, such as the Bcr-Abl oncoprotein p210 that is characteristic of chronic myelogenous leukaemia (CML), has provided excellent proof of principle for RNAi as an anti-cancer therapeutic agent. For CML, the main treatment options have been chemotherapy, allogeneic bone-marrow transplant and most recently, the use of a small molecule, the tyrosine kinase inhibitor, imatinib. Despite initial excitement about the potential of imatinib, a growing number of patients have developed resistance to it^{86–89}, necessitating alternative forms of therapy. Bcr-Abl p210 has been selectively downregulated by both synthetic siRNAs and lentiviral-vector-transduced shRNAs in cell lines^{90–92}. Importantly, the downregulation is selective for only the p210 oncoprotein and its mRNA, which results in inhibition of cell proliferation as a direct consequence of RNAi. Haematologic malignancies are often treated by bone-marrow transplantation. Therefore, a possible therapeutic application would be to transduce haematopoietic stem cells with vectors harbouring a gene that targets the mRNA encoding the oncogenic p210 protein, thereby protecting patients from relapse caused by proliferation of latent leukaemic stem cells. Again, delivery is the key issue; 100% transduction of the stem cells reinfused in a bone-marrow transplant setting will be required to make this therapeutically effective. The improvements in viral vector titres and transduction efficiencies may make this possible.

RNAi for genetic diseases

A promising lead towards using RNAi for the treatment of genetic diseases has been provided by preliminary studies that demonstrate how SNPs in mutant allele transcripts can be used as selective targets for RNAi^{93,94}. Finding an siRNA that is highly selective for a particular SNP is a challenge, but has been accomplished by systematic analyses of siRNAs in which the polymorphic nucleotide is complementary to the mid-region of the siRNA. In certain examples, the siRNAs direct selective degradation of only the mutant transcripts, leaving the wild-type transcripts intact despite only a single mismatch^{93,94}. Another example of siRNAs targeting an SNP was recently reported in studies of amyotrophic lateral sclerosis (ALS) caused by mutations in the Cu, Zn superoxide dismutase (*SOD1*) gene⁹⁵. Because the wild-type *SOD1* performs important functions, it is important to selectively eliminate expression of only the mutant allelic transcript. Many *SOD1* mutations are single-nucleotide changes. Ding *et al.*⁹⁵ achieved selective degradation of a mutant *SOD1* allele, thereby providing a potential therapeutic application for the treatment of ALS.

Disease-causing polyglutamine proteins encoded by CAG-repeat-containing transcripts are found in several neurological diseases such as Huntington's disease. These proteins are especially challenging targets for RNAi because CAG repeats are common to many normal transcripts as well, and the repeats themselves cannot be selectively targeted by siRNAs. But with the recent finding that delivery of siRNAs and viral vectors expressing siRNAs to diseased regions of the brain is technically feasible⁹⁶, coupled with selective targeting of SNPs in the mutant transcripts, the promise of clinical use of RNAi for the treatment of degenerative, neurological diseases should be realised.

Challenges for RNAi as a therapy

Two key challenges in developing RNAi as a therapy are avoiding off-target effects and ensuring efficient delivery. One potential risk for side effects emerges from the feature that distinguishes RNAi from other antisense technologies — the use of cellular machinery for directing sequence-specific silencing. This machinery has specific purposes, such as miRNA-mediated gene regulation^{97,98}. Using siRNAs to target specific cellular or viral transcripts in essence hijacks the endogenous RNAi machinery, and we know little about the potential for saturating the RNAi pathway in primary cells, although saturation of RISC is demonstrable in cultured cells⁹⁹. So endogenous RNAi pathways could potentially be affected by siRNAs. It will be important to pay close attention to basic research studies on off-target effects of siRNAs and on the design of effective siRNAs^{22,27,100}.

A better understanding of the mechanisms that lead to nonspecific effects of short dsRNAs is essential before the use of siRNAs or shRNAs can be tested in patient trials.

The issue of delivery has restricted the antisense field for almost two decades. It is feasible to infuse backbone-modified oligonucleotides *in vivo*, but achieving intracellular delivery at therapeutically effective concentrations is a major challenge. Targeted delivery to specific cell or tissue types is still not a practical reality for oligonucleotide-based therapeutics. The alternative approach is viral-vector-mediated delivery of therapeutic shRNA genes. Because this type of delivery results in gene therapy, there are several associated safety concerns, and systemic delivery of viral vectors is still a major hurdle. Nevertheless, the potency and potential general therapeutic utility of RNAi is prompting renewed vigour into delivery-related research. It remains to be determined whether backbone-modified, nuclease-resistant siRNAs will move to the clinic more quickly than synthetic deoxyoligonucleotides.

Perspective

In a remarkably short time since its discovery in model organisms, the RNAi pathway has emerged as a powerful tool for the study of gene function in mammals. As our understanding of the under-lying biology and biochemistry of this conserved gene-regulatory mechanism improves, so does our ability to exploit RNAi as an experimental tool. With the use of RNAi in whole animals increasing, we anticipate growing enthusiasm for the use of RNAi triggers in therapy. Despite considerable hurdles to overcome, it seems likely that RNAi will find a place alongside more conventional approaches in the treatment of diseases, although it is unclear how long we will have to wait to witness the first RNAi-based drug. The big question is whether RNAi can revolutionize the treatment of human disease in the same way that it has revolutionized basic research into gene function. □

doi:10.1038/nature02870□

1. Fire, A., Albertson, D., Harrison, S. W. & Moerman, D. G. Production of antisense RNA leads to effective and specific inhibition of gene expression in *C. elegans* muscle. *Development* **113**, 503–514 (1991).
2. Guo, S. & Kemphues, K. J. *par-1*, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* **81**, 611–620 (1995).
3. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
4. Hannon, G. J. RNA interference. *Nature* **418**, 244–251 (2002).
5. Williams, B. R. Role of the double-stranded RNA-activated protein kinase (PKR) in cell regulation. *Biochem. Soc. Trans.* **25**, 509–513 (1997).
6. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
7. Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R. & Hannon, G. J. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**, 1146–1150 (2001).
8. Elbashir, S. M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).
9. Hutvagner, G. *et al.* A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. *Science* **293**, 834–838 (2001).
10. Grishok, A. *et al.* Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106**, 23–34 (2001).
11. Ketting, R. F. *et al.* Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* **15**, 2654–2659 (2001).
12. Padison, P. J., Caudy, A. A., Sachidanandam, R. & Hannon, G. J. Short hairpin activated gene silencing in mammalian cells. *Methods Mol. Biol.* **265**, 85–100 (2004).
13. Khvorova, A., Reynolds, A. & Jayasena, S. D. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209–216 (2003).
14. Schwarz, D. S. *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199–208 (2003).
15. Reynolds, A. *et al.* Rational siRNA design for RNA interference. *Nature Biotechnol.* **22**, 326–330 (2004).
16. Silva, J. M., Sachidanandam, R. & Hannon, G. J. Free energy lights the path toward more effective RNAi. *Nature Genet.* **35**, 303–305 (2003).
17. Aza-Blanc, P. *et al.* Identification of modulators of TRAIL-induced apoptosis via RNAi-based phenotypic screening. *Mol. Cell* **12**, 627–637 (2003).
18. Pebernard, S. & Iggo, R. D. Determinants of interferon-stimulated gene induction by RNAi vectors. *Differentiation* **72**, 103–111 (2004).
19. Persengiev, S. P., Zhu, X. & Green, M. R. Nonspecific, concentration-dependent stimulation and repression of mammalian gene expression by small interfering RNAs (siRNAs). *RNA* **10**, 12–18 (2004).
20. Sledz, C. A., Holko, M., de Veer, M. J., Silverman, R. H. & Williams, B. R. Activation of the interferon system by short-interfering RNAs. *Nature Cell Biol.* **5**, 834–839 (2003).
21. Kim, D. H. *et al.* Interferon induction by siRNAs and ssRNAs synthesized by phage polymerase. *Nature Biotechnol.* **22**, 321–325 (2004).
22. Jackson, A. L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nature Biotechnol.* **21**, 635–637 (2003).
23. Doench, J. G. & Sharp, P. A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504–511 (2004).

24. Lai, E. C., Tomancak, P., Williams, R. W. & Rubin, G. M. Computational identification of *Drosophila* microRNA genes. *Genome Biol.* **4**, R42 (2003).
25. Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115**, 787–798 (2003).
26. Enright, A. J. *et al.* MicroRNA targets in *Drosophila*. *Genome Biol.* **5**, R1 (2003).
27. Scacheri, P. C. *et al.* Short interfering RNAs can induce unexpected and divergent changes in the levels of untargeted proteins in mammalian cells. *Proc. Natl Acad. Sci. USA* **101**, 1892–1897 (2004).
28. McCaffrey, A. P. *et al.* RNA interference in adult mice. *Nature* **418**, 38–39 (2002).
29. Matsuda, T. & Cepko, C. L. Electroporation and RNA interference in the rodent retina *in vivo* and *in vitro*. *Proc. Natl Acad. Sci. USA* **101**, 16–22 (2004).
30. Layzer, J. M. *et al.* *In vivo* activity of nuclease-resistant siRNAs. *RNA* **10**, 766–771 (2004).
31. Lewis, D. L., Hagstrom, J. E., Loomis, A. G., Wolff, J. A. & Herweijer, H. Efficient delivery of siRNA for inhibition of gene expression in postnatal mice. *Nature Genet.* **32**, 107–108 (2002).
32. Sorensen, D. R., Leirdal, M. & Sioud, M. Gene silencing by systemic delivery of synthetic siRNAs in adult mice. *J. Mol. Biol.* **327**, 761–766 (2003).
33. Brummelkamp, T. R., Bernards, R. & Agami, R. Stable suppression of tumorigenicity by virus-mediated RNA interference. *Cancer Cell* **2**, 243–247 (2002).
34. Robinson, D. A. *et al.* A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. *Nature Genet.* **33**, 401–406 (2003).
35. Hemann, M. T. *et al.* An epi-allelic series of p53 hypomorphs created by stable RNAi produces distinct tumor phenotypes *in vivo*. *Nature Genet.* **33**, 396–400 (2003).
36. Carmell, M. A., Zhang, L., Konkin, D. S., Hannon, G. J. & Rosenquist, T. A. Germline transmission of RNAi in mice. *Nature Struct. Biol.* **10**, 91–92 (2003).
37. Tiscornia, G., Singer, O., Ikawa, M. & Verma, I. M. A general method for gene knockdown in mice by using lentiviral vectors expressing small interfering RNA. *Proc. Natl Acad. Sci. USA* **100**, 1844–1848 (2003).
38. Hasuwa, H., Kaseda, K., Einarsdottir, T. & Okabe, M. Small interfering RNA and gene silencing in transgenic mice and rats. *FEBS Lett.* **532**, 227–230 (2002).
39. Kunath, T. *et al.* Transgenic RNA interference in ES cell-derived embryos recapitulates a genetic null phenotype. *Nature Biotechnol.* **21**, 559–561 (2003).
40. Hsieh, A. C. *et al.* A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens. *Nucleic Acids Res.* **32**, 893–901 (2004).
41. Brummelkamp, T. R., Nijman, S. M., Dirac, A. M. & Bernards, R. Loss of the cyclinomatosis tumour suppressor inhibits apoptosis by activating NF- κ B. *Nature* **424**, 797–801 (2003).
42. Berns, K. *et al.* A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431–437 (2004).
43. Paddison, P. J. *et al.* A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**, 427–431 (2004).
44. Zheng, L. *et al.* An approach to genomewide screens of expressed small interfering RNAs in mammalian cells. *Proc. Natl Acad. Sci. USA* **101**, 135–140 (2004).
45. Luo, B., Heard, A. D. & Lodish, H. F. Small interfering RNA production by enzymatic engineering of DNA (SPEED). *Proc. Natl Acad. Sci. USA* **101**, 5494–5499 (2004).
46. Shirane, D. *et al.* Enzymatic production of RNAi libraries from cDNAs. *Nature Genet.* **36**, 190–196 (2004).
47. Sen, G., Wehrman, T. S., Myers, J. W. & Blau, H. M. Restriction enzyme-generated siRNA (REGS) vectors and libraries. *Nature Genet.* **36**, 183–189 (2004).
48. Silva, J. M., Mizuno, H., Brady, A., Lucito, R. & Hannon, G. J. RNA interference microarrays: high-throughput loss-of-function genetics in mammalian cells. *Proc. Natl Acad. Sci. USA* **101**, 6548–6552 (2004).
49. Ziauddin, J. & Sabatini, D. M. Microarrays of cells expressing defined cDNAs. *Nature* **411**, 107–110 (2001).
50. Jacque, J. M., Triques, K. & Stevenson, M. Modulation of HIV-1 replication by RNA interference. *Nature* **418**, 435–438 (2002).
51. Lee, N. S. *et al.* Expression of small interfering RNAs targeted against HIV-1 rev transcripts in human cells. *Nature Biotechnol.* **20**, 500–505 (2002).
52. Coburn, G. A. & Cullen, B. R. Potent and specific inhibition of human immunodeficiency virus type-1 replication by RNA interference. *J. Virol.* **76**, 9225–9231 (2002).
53. Surabhi, R. M. & Gaynor, R. B. RNA interference directed against viral and cellular targets inhibits human immunodeficiency virus type-1 replication. *J. Virol.* **76**, 12963–12973 (2002).
54. Novina, C. D. *et al.* siRNA-directed inhibition of HIV-1 infection. *Nature Med.* **8**, 681–686 (2002).
55. Park, W. S. *et al.* Prevention of HIV-1 infection in human peripheral blood mononuclear cells by specific RNA interference. *Nucleic Acids Res.* **30**, 4830–4835 (2002).
56. Boden, D., Pusch, O., Lee, F., Tucker, L. & Ramratnam, B. Human immunodeficiency virus type-1 escape from RNA interference. *J. Virol.* **77**, 11531–11535 (2003).
57. Martinez, M. A. *et al.* Suppression of chemokine receptor expression by RNA interference allows for inhibition of HIV-1 replication. *AIDS* **16**, 2385–2390 (2002).
58. Capodici, J., Kariko, K. & Weissman, D. Inhibition of HIV-1 infection by small interfering RNA-mediated RNA interference. *J. Immunol.* **169**, 5196–5201 (2002).
59. Banerjee, A. *et al.* Inhibition of HIV-1 by lentiviral vector-transduced siRNAs in T lymphocytes differentiated in SCID-hu mice and CD34⁺ progenitor cell-derived macrophages. *Mol. Ther.* **8**, 62–71 (2003).
60. Li, M. J. *et al.* Inhibition of HIV-1 infection by lentiviral vectors expressing Pol III-promoted anti-HIV RNAs. *Mol. Ther.* **8**, 196–206 (2003).
61. Eugen-Olsen, J. *et al.* Heterozygosity for a deletion in the *CCR-5* gene leads to prolonged AIDS-free survival and slower CD4 T-cell decline in a cohort of HIV-seropositive individuals. *AIDS* **11**, 305–310 (1997).
62. Samson, M. *et al.* Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the *CCR-5* chemokine receptor gene. *Nature* **382**, 722–725 (1996).
63. Qin, X. F., An, D. S., Chen, I. S. & Baltimore, D. Inhibiting HIV-1 infection in human T cells by lentiviral-mediated delivery of small interfering RNA against CCR5. *Proc. Natl Acad. Sci. USA* **100**, 183–188 (2003).
64. Dell'Agnola, C. *et al.* *In vitro* and *in vivo* hematopoietic potential of human stem cells residing in muscle tissue. *Exp. Hematol.* **30**, 905–914 (2002).
65. Dropulic, B. Lentivirus in the clinic. *Mol. Ther.* **4**, 511–512 (2001).
66. Davis, B. M., Humeau, L. & Dropulic, B. *In vivo* selection for human and murine hematopoietic cells transduced with a therapeutic MGMT lentiviral vector that inhibits HIV replication. *Mol. Ther.* **9**, 160–172 (2004).
67. Amado, R. G. *et al.* Anti-human immunodeficiency virus hematopoietic progenitor cell-delivered ribozyme in a phase I study: myeloid and lymphoid reconstitution in human immunodeficiency virus type-1-infected patients. *Hum. Gene Ther.* **15**, 251–262 (2004).
68. Michienzi, A. *et al.* RNA-mediated inhibition of HIV in a gene therapy setting. *Ann. NY Acad. Sci.* **1002**, 63–71 (2003).
69. McCaffrey, A. P. *et al.* Inhibition of hepatitis B virus in mice by RNA interference. *Nature Biotechnol.* **6**, 639–644 (2003).
70. Blight, K. J., Kolykhalov, A. A. & Rice, C. M. Efficient initiation of HCV RNA replication in cell culture. *Science* **290**, 1972–1974 (2000).
71. Lohmann, V. *et al.* Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science* **285**, 110–113 (1999).
72. Ikeda, M., Yi, M., Li, K. & Lemon, S. M. Selectable subgenomic and genome-length dicistronic RNAs derived from an infectious molecular clone of the HCV-N strain of hepatitis C virus replicate efficiently in cultured Huh7 cells. *J. Virol.* **76**, 2997–3006 (2002).
73. Pietschmann, T., Lohmann, V., Rutter, G., Kurpanek, K. & Bartenschlager, R. Characterization of cell lines carrying self-replicating hepatitis C virus RNAs. *J. Virol.* **75**, 1252–1264 (2001).
74. Randall, G., Grakoui, A. & Rice, C. M. Clearance of replicating hepatitis C virus replicon RNAs in cell culture by small interfering RNAs. *Proc. Natl Acad. Sci. USA* **100**, 235–240 (2003).
75. Wilson, J. A. *et al.* RNA interference blocks gene expression and RNA synthesis from hepatitis C replicons propagated in human liver cells. *Proc. Natl Acad. Sci. USA* **100**, 2783–2788 (2003).
76. Kapadia, S. B., Brideau-Andersen, A. & Chisari, F. V. Interference of hepatitis C virus RNA replication by short interfering RNAs. *Proc. Natl Acad. Sci. USA* **100**, 2014–2018 (2003).
77. Song, E. *et al.* RNA interference targeting Fas protects mice from fulminant hepatitis. *Nature Med.* **9**, 347–351 (2003).
78. Eastman, S. J. *et al.* Development of catheter-based procedures for transducing the isolated rabbit liver with plasmid DNA. *Hum. Gene Ther.* **13**, 2065–2077 (2002).
79. Kittler, R. & Buchholz, F. RNA interference: gene silencing in the fast lane. *Semin. Cancer Biol.* **13**, 259–265 (2003).
80. Wall, N. R. & Shi, Y. Small RNA: can RNA interference be exploited for therapy? *Lancet* **362**, 1401–1403 (2003).
81. Lu, P. Y., Xie, F. Y. & Woodle, M. C. siRNA-mediated antitumorigenesis for drug target validation and therapeutics. *Curr. Opin. Mol. Ther.* **5**, 225–234 (2003).
82. Buche, T. Proapoptotic therapy with oblimersen (bcl-2 antisense oligonucleotide)—review of preclinical and clinical results. *Onkologie* **26** (Suppl. 7), 60–69 (2003).
83. Chiu, Y. L. & Rana, T. M. siRNA function in RNAi: a chemical modification analysis. *RNA* **9**, 1034–1048 (2003).
84. Czauderna, F. *et al.* Structural variations and stabilising modifications of synthetic siRNAs in mammalian cells. *Nucleic Acids Res.* **31**, 2705–2716 (2003).
85. Wang, L., Prakash, R. K., Stein, C. A., Koehn, R. K. & Ruffner, D. E. Progress in the delivery of therapeutic oligonucleotides: organ/cellular distribution and targeted delivery of oligonucleotides *in vivo*. *Antisense Nucleic Acid Drug Dev.* **13**, 169–189 (2003).
86. Holtz, M. S. & Bhatia, R. Effect of imatinib mesylate on chronic myelogenous leukemia hematopoietic progenitor cells. *Leuk. Lymphoma* **45**, 237–245 (2004).
87. Tauchi, T. & Ohayashiki, K. Molecular mechanisms of resistance of leukemia to imatinib mesylate. *Leuk. Res.* **28** (Suppl. 1), 39–45 (2004).
88. Cowan-Jacob, S. W. *et al.* Imatinib (STI571) resistance in chronic myelogenous leukemia: molecular basis of the underlying mechanisms and potential strategies for treatment. *Mini Rev. Med. Chem.* **4**, 285–299 (2004).
89. Marcucci, G., Perrotti, D. & Caligiuri, M. A. Understanding the molecular basis of imatinib mesylate therapy in chronic myelogenous leukemia and the related mechanisms of resistance. Commentary. *Clin. Cancer Res.* **9**, 1333–1337, 2003. *Clin. Cancer Res.* **9**, 1248–1252 (2003).
90. Li, M. J., McMahon, R., Snyder, D. S., Yee, J. K. & Rossi, J. J. Specific killing of Ph⁺ chronic myeloid leukemia cells by a lentiviral vector-delivered anti-bcr/abl small hairpin RNA. *Oligonucleotides* **13**, 401–409 (2003).
91. Wohlbold, L. *et al.* Inhibition of *bcr-abl* gene expression by small interfering RNA sensitizes for imatinib mesylate (STI571). *Blood* **102**, 2236–2239 (2003).
92. Scherr, M. *et al.* Specific inhibition of *bcr-abl* gene expression by small interfering RNA. *Blood* **101**, 1566–1569 (2003).
93. Miller, V. M., Gouvion, C. M., Davidson, B. L. & Paulson, H. L. Targeting Alzheimer's disease genes with RNA interference: an efficient strategy for silencing mutant alleles. *Nucleic Acids Res.* **32**, 661–668 (2004).
94. Miller, V. M. *et al.* Allele-specific silencing of dominant disease genes. *Proc. Natl Acad. Sci. USA* **100**, 7195–7200 (2003).
95. Ding, H. *et al.* Selective silencing by RNAi of a dominant allele that causes amyotrophic lateral sclerosis. *Aging Cell* **2**, 209–217 (2003).
96. Davidson, B. L. & Paulson, H. L. Molecular medicine for the brain: silencing of disease genes with RNA interference. *Lancet Neurol.* **3**, 145–149 (2004).
97. Pasquini, A. E. MicroRNAs: deviants no longer. *Trends Genet.* **18**, 171–173 (2002).
98. Moss, E. G. MicroRNAs: hidden in the genome. *Curr. Biol.* **12**, R138–R140 (2002).
99. Hutvagner, G., Simard, M. J., Mello, C. C. & Zamore, P. D. Sequence-specific inhibition of small RNA function. *PLoS Biol.* **2**, E98 (2004).
100. Saxena, S., Jonsson, Z. O. & Dutta, A. Small RNAs with imperfect match to endogenous mRNA repress translation. Implications for off-target activity of small inhibitory RNA in mammalian cells. *J. Biol. Chem.* **278**, 44312–44319 (2003).

Acknowledgements We thank members of the RNAi community for making this an exciting field of research. G.J.H. thanks F. Rivas for comments on the manuscript and J. Duffy for help with the figures. J.J.R. thanks L. Scherer for suggestions on therapeutic applications of RNAi. G.J.H. is supported by an Innovator award from the US Army Breast Cancer Research Program and by grants from the NIH. J.J.R. is supported by grants from the NIH, NIAID and NHBLI.

Competing interests statement The authors declare that they have no competing financial interests.

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

US Head Office, New York

345 Park Avenue South, 10th Floor,
New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager/ Corporations:

Peter Bless

Classified Sales Representatives:

Phone: +1 800 989 7718

**Pennsylvania, New York,
Maryland/ Latin America/ NIH:**

Michael McGillion

Midwest USA:

Wade Tucker

East USA/ Canada:

Janine Taormina

San Francisco Office

Classified Sales Representative:

Michaela Bjorkman

West USA/ West Corp. Canada

225 Bush Street, Suite 1453

San Francisco, CA 94104

Tel +1 415 781 3803

Fax +1 415 781 3805

e-mail: m.bjorkman@nature.sf.com

European Head Office, London

The Macmillan Building, 4 Crinan Street,
London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayoumi (4978)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Stefan Hales

European Satellite Office

Germany/ Austria/ Italy/

The Netherlands/ Belgium:

Patrick Phelan, Odo Wulffien

e-mail: p.phelan@nature.com

o.wulffien@nature.com

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Harakatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: rasami@naturejpn.com

naturejobs

Looking for a fast track

If you are a young scientist, time is not on your side. In the United States and many parts of Europe, the average length of PhD study has been increasing. Meanwhile, many young scientists have found it necessary to do multiple postdocs in order to gain enough publications to secure a professorship. Making that next step, to independent investigator, results in the biggest lag of all. Following postdoctoral fellowships, many scientists end up taking one short-term position after another in the hope of eventually landing permanent employment.

These issues have at last caught the attention of policy-makers. In the European Union (EU), the 'Bologna process' aims to standardize PhDs into four years across its member states, but the countries remain unsure how to meet this goal. In addition, researchers in Britain are worried that an EU rule on the length of temporary contracts might actually jeopardize job security (see *Nature* 431, 6; 2004). And in the United States, the National Academy of Sciences has done a good job of reporting on the plight of the postdoc, but it has yet to address the length and number of fellowships young scientists must take to succeed.

Of course, it is easy to identify problems — it is much harder to present solutions. So, over the next two months, *Naturejobs* will attempt to do both in its 'Fast Track' series. The first instalment, published this issue, takes a look at PhDs (see overleaf). As well as examining the general trends and challenges faced by young scientists, the series will highlight examples of scientists who have swiftly navigated the various stages of their career.

Even if institutions do go ahead and reform each stage of the scientific career process, young scientists might do well to look to these examples and plot their careers accordingly. Institutions might place themselves on the fast track eventually, but today's young scientists don't have time to wait.

Paul Smaglik
Naturejobs editor



Contents

SPECIAL REPORT

The changing length
of PhDs p382

CAREER VIEW

Recruiters & Industry
Building an intern programme
Graduate Journal
Tunnel vision
Movers
Tim Morley p384

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Four years has become the magic number for many graduate programmes in the United States and Europe. Eugene Russo explains the logic behind the maths.

As a prospective graduate student five years ago, Amy Caudy was interested in the new Watson School of Biological Sciences at Cold Spring Harbor Laboratory (CSHL) in New York state. Her reasons: CSHL's reputation, the small class size and an innovative curriculum. The school's insistence that students complete their PhDs within four years, unusually fast for life-sciences students in the United States, wasn't much of a factor. "But at a certain point in graduate school it became a major factor," Caudy chuckles.

US graduate programmes, especially in life sciences, have grown excruciatingly long in recent decades. Eight years is not unusual. According to Jim Voytuk, senior programme officer at the US National Academy of Sciences, the average time to PhD since the 1970s has increased by nearly a year in chemistry and physics and nearly 18 months in the biosciences. In 1973–82, the average in biosciences was 6.3 years; for 1993–2002, the average was 7.7 years. In physics, time to degree increased from 6.6 to 7.4 years, in chemistry from 5.8 to 6.7 years.

In the United Kingdom, the trend is up from three years to three-and-a-half or four. With financial commitments from the Medical Research Council (MRC) and the Wellcome Trust, the number of four-year programmes is rising. The rest of Europe (where students spend longer than Britain's standard three years as undergraduates) is following suit. The

'Bologna process', led by the European Union, aims to standardize graduate studies by 2010, with four-year doctoral programmes as one aim, according to Iain Cameron, head of postgraduate training at the UK Engineering and Physical Sciences Research Council.

Ironically, the new European ideas are modelled on the US system, where four- or five-year programmes often stretch much longer. The way the Watson School's first six students gained PhDs in an average of four years may provide lessons for other institutions. The lengthening in Britain, meanwhile, has provided a model in which students are given more freedom to explore different lab environments and to research fresh disciplines.

Some people suggest that the entire area needs to be rethought. "It's important that the graduate and postdoctoral period are thought about as a combined training period, which is often not the case," says James William Nelson, a professor of cellular physiology at Stanford Medical School and former senior associate



Amy Caudy: the four-year PhD programme works "because people are beating on you to drag you out the door and get you to progress".

dean of graduate and postdoctoral education.

Students need to consider where their aspirations lie, Nelson suggests. If academia is not the ultimate objective, doing a postdoc should not be a given; nor should staying on at graduate school, waiting to get a paper in *Nature* or *Science*.

SHARPER FOCUS ON CAREER

Many programmes are working to shorten the early stages, such as courses and oral tests, says Keith Yamamoto, vice-dean for research at the medical school of the University of California, San Francisco. But it is the later research that takes so much time. "If we can instil a sharper focus on what they want to be coming out of graduate school with, I think people will discover that they can acquire those tools in four years," says Yamamoto.

Science students should think about their careers the way business students think about theirs, says Raymond Clark, co-chair of the policy committee for the US National Postdoctoral Association. "Look to leverage the greatest advantage in your direction," he says. "It's your career you need to be concerned about, not your mentor's." If your mentor doesn't support you, you need to find a new one, Clark advises (see *Nature* 422, 784–785; 2003).

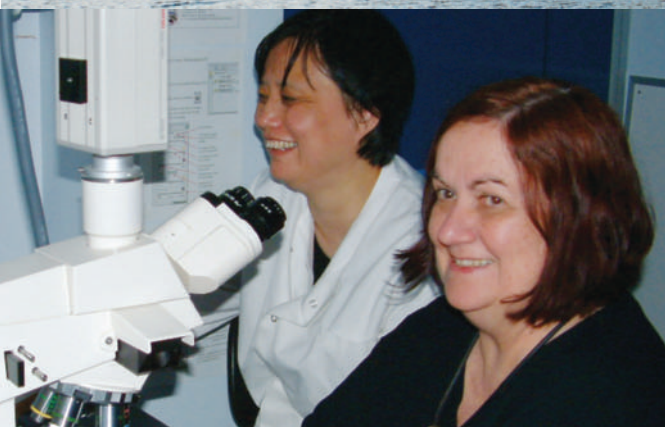
Nelson says that graduate students should keep three aims in mind: scholarship, or being able to define a problem; training as experimentalists; and critical thinking skills to interpret data and the literature.

Finding the right lab is crucial to accomplishing all





COLD SPRING HARBOR LAB



Anne Mudge (with colleague Lili Cheng, left): lab rotation system works better than matching students to supervisors. Above, Cold Spring Harbor.

four years because it takes only a few students a year, and offers structured supervision so that they don't fall behind. Replicating that environment at larger universities may be difficult. Also, Watson's small class size can mean sacrificing some options; students' curricula are pretty well set. For instance, there is no maths department, so a student interested in statistics would have to either go without or look elsewhere.

Clark encourages students hoping for jobs and for fairly short graduate and postdoc experiences to opt for cutting-edge fields such as bioinformatics, nanotechnology or biodefence. Caudy, now a postdoc at Washington University in St Louis, got lucky: she chose RNA interference (RNAi). She started graduate research in 1999, just as the technique exploded on the scene, and she contributed to key RNAi papers.

PRESSURE TO PUBLISH

But shorter study means fewer publications, a concern for students in the job market. Mudge recognizes the difficulty of getting a solid piece of research done within four years. "Everything has to go well," she says.

In Europe, hanging around a lab for years is hard because the funding isn't there. And in some countries, other requirements complicate and lengthen the PhD process. In Finland, for example, PhD students must first complete a master's, which pushes the average time to a PhD to nearly eight years. Some Finnish scientists opt instead to do MD/PhDs even though they don't intend to practise medicine, because they can begin thesis research earlier than in an MSc. Students, postdocs and faculty members at a planning retreat for the University of Oulu's Biocenter last month agreed that four years is a good goal for a PhD — but wondered whether it is attainable, given the growing pressure to complete a PhD with more and better publications.

"Time will have to tell if it'd be better to stick around for another six months and have that next paper," says Caudy, who has segued into immunology as a postdoc at Washington University. "It'll be interesting to see in three to four years from now where people end up."

Eugene Russo is a freelance science writer in Takoma Park, Maryland.

three. At the Watson School, Caudy and her classmates went through three six-week lab rotations, a crucial facet of the educational experience at many other institutions, including the European Molecular Biology Laboratory in Heidelberg. Importantly, she says, the rotations were not open-ended, as in many programmes. Students can't fall into the trap of thinking that if they just stay a few more weeks, they will complete a task — then find themselves in the same place a year later. "The school works because of people like the dean, Lilian Gann, who are beating on you to drag you out the door at the end of your six weeks and get you to progress," says Caudy.

At University College London's four-year MRC-funded PhD programme in molecular cell biology, started in 1993, programme director Anne Mudge cites newly instituted lab rotations as a major improvement. In the previous system, studentships were allocated to supervisors, who were then matched up with students. "Students were sort of obliged to go to the person who had the studentship as opposed to being driven by interests," says Mudge. "Sometimes that matched up well, but sometimes it didn't." Programme graduate Nic Tapon, now at the charity Cancer Research UK, says the rotations provided insight into lab dynamics, and gave some sense of where he would be happy.

The Watson School pushes students, providing both an academic and a research adviser. Crucially, say Caudy and Gann, it is the school, not the student, that schedules thesis committee meetings with busy faculty members. "It saves time and aggravation," says Gann.

In part, though, the Watson School is able to stick to

GRADUATE JOURNAL

Tunnel vision

It's the last train home and I'm scribbling corrections on my particle-physics paper. Out of the corner of my eye I notice the man next to me looking over my shoulder at my work. It's late, he's probably bored, I tell myself. But does he really care what I do? Should he care? Do I care whether he cares?

Yes! If nothing else, his taxes contribute to the millions needed to build the high-tech detectors used in my area of physics. They pay for me to go to conferences, for my computer and for the hours I sit at the screen fixing broken code.

Scientists have a duty to put something back into the system that brought us this far. On a more practical level, if we don't sell our work to the politicians who hand out the taxes, there will be no more research.

But what's the best way to reach out? Expounding the merits of electroweak symmetry breaking at the supermarket checkout is probably not the best approach. You need something people can relate to — keep it simple and tell a story.

The curious stranger was a reminder not to become too blinkered. Scientists can't operate in a bubble. We need to share our excitement for what we do with different audiences. The rest of the world is eager to listen, we just have to speak to it in the right way. ■

Amber Jenkins is a graduate student in particle physics at Imperial College, London.

RECRUITERS & INDUSTRY

Building an intern programme

IBM may be viewed by many as a giant in information technology, but it hasn't always been the first choice for jobseekers. As the Internet boomed in the 1990s, young people who wanted to work at the frontiers of computer science saw IBM as irrelevant, says Jane Harper, who runs the company's programme for talent-scouting at universities. Instead, the brightest and best joined Internet start-ups, where they felt they would have more direct involvement in developing new technologies.

To arrest this trend, IBM set up a US intern programme in 1996 called Extreme Blue (derived from the firm's nickname 'Big Blue'). Aimed at undergraduates, master's and PhD students one year away from finishing their degree, Extreme Blue puts

the interns at the cutting edge by assigning them to projects that have a genuine chance of reaching the market-place.

Under the scheme, three scientific and technical interns are teamed with an MBA student for the commercial angle, and the group is then guided by an IBM mentor. The 12-week internships have proved so popular, says Harper, that this year nearly 4,500 students applied for the 200 available places.

Extreme Blue is part of a broad strategy at IBM, which this year will see the company take on some 2,000 interns in the United States alone. Other schemes include ProjectView, which targets women, under-represented minorities and people with disabilities.

Internships make for a good trial period for both employer and prospective employees, notes Harper. Indeed, IBM sees its various intern programmes as a good way to help it fill

its employment pipeline. At the beginning of this year, the company announced plans to expand its workforce by some 15,000 people. Many of these new jobs are likely to be based in China or India, so it is no surprise that Extreme Blue has now been implemented in both Beijing and Bangalore — as well as in Toronto, Canada. The programme is also being run year-round rather than only during the summer.

But perhaps the true measure of Extreme Blue's success, says Harper, is how many of the programme's interns stay on to become full-time employees. Even though the dotcom bubble has well and truly burst, "it's just as hard to get the top talent", she says. Nevertheless, she adds, Extreme Blue is attracting — and retaining — a lot more top talent than IBM was seeing ten years ago. ■

Paul Smaglik is editor of Naturejobs.

MOVERS

Tim Morley, vice-president, preclinical sciences, Ardana, Edinburgh, Scotland



Tim Morley was still an undergraduate when he got his first taste of industry — but his brief spell in a commercial lab set the direction of his future career.

Morley was doing a 'sandwich' degree in biochemistry at the University of Liverpool, UK, which meant that he got to spend some of his course doing an internship. At the suggestion of one of his tutors, he did his placement

in southeast England at the drug company Beecham.

The experience was a real eye-opener, says Morley, as he had never seen chemistry done on such a large scale, with the latest equipment and instrumentation. "As an undergraduate, you have limited exposure to state-of-the-art kit," he says.

His time at Beecham helped him to refine the direction of his studies, and he went on to focus on pharmacology, toxicology, drug metabolism and pharmacokinetics. The strategy paid off, landing him a job at Wellcome Research Laboratories in Beckenham before he did his PhD. After nine years there he moved briefly to Agrevo before joining Vernalis Research in Wokingham.

Morley still believes that drug metabolism and toxicology are good areas for a scientific career. In the course of his professional life, he notes, he has

seen a consistent shortage of people with the right skills for this kind of work — particularly scientists who can create and evaluate drugs.

In his latest move, Morley has accepted the post of vice-president, preclinical sciences, at Ardana, a small drug company in Edinburgh. The firm was spun off in 2000 from the Human Reproductive Sciences Unit run by the UK Medical Research Council (MRC). Morley will coordinate the preclinical evaluation of the company's drugs, and will liaise with MRC staff in Edinburgh and at St Mary's Hospital for Women and Children in Manchester. He is currently searching for other centres of excellence in human reproductive health for potential collaborations.

Morley says that he is looking forward to his new role. "If you have a wealth of experiences," he says, "you can make a bigger impact in a smaller company." ■

CV

1997–2004: Senior director of biological research, Vernalis Research, Wokingham, UK

1996–97: Senior scientist in toxicology, Agrevo (now Aventis), Chesterford Park, UK

1988–96: Senior scientist in drug-safety evaluation, Wellcome Research Laboratories, Beckenham, UK

1993–97: PhD in mechanistic toxicology, University of Aberdeen, Scotland

1990–91: MSc in biopharmacy, King's College Chelsea, London, UK